Title

Levels of Processing in the Pathways for Intelligible Speech

Author

Maddox, Christopher Dale

Publication Date

2015

UNIVERSITY OF CALIFORNIA, IRVINE

Levels of Processing in the Pathways for Intelligible Speech

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Cognitive Neuroscience

by

Christopher Dale Maddox, Jr.

Dissertation Committee:
Professor Gregory Hickok, Chair
Professor Kourosh Saberi
Professor Emily Grossman

2015

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

First and foremost I would like to express my gratitude to Dr. Gregory S. Hickok. This manuscript is in large part a reflection of his expertise and patience. I am grateful for the opportunity he gave me to be a productive member of the scientific community, as well as all the support I have received over the years.

I would also like to thank all members, past and present, of the Hickok Lab (Auditory and Language Neuroscience Lab, formerly Lab for Cognitive Brain Research). It was the best work situation a graduate student can expect to find. I would particularly like to acknowledge Dr. Kayoka Okada, a contributor to the projects discussed in this manuscript and early mentor of mine. Dr. Kourosh Saberi also made a considerable contribution to our research.

I would also like to thank Dr. Emily Grossman, the third member of my advisory committee (joining Hickok and Saberi).

The National Sciences Foundation's Graduate Research Fellowships Program was of great benefit to myself and our research. The material here is based upon work supported by the NSF under Grant No. 1000106691.

Finally, I would like to thank the University of California, Irvine. It is a wonderful institution with a substantive focus on post-graduate development.

CURRICULUM VITAE

Christopher Dale Maddox

2007	B.S. in Psychology, University of Alabama
	B.S. in Philosophy, University of Alabama
	B.S. in Computer Science, University of Alabama
2015	Ph.D. in Cognitive Neuroscience, University of California, Irvine, Department of Cognitive Sciences

FIELD OF STUDY

Neurolinguistic Processing in Temporal Cortex

Research Experience

August 2008 to present

Graduate Student Researcher for Dr. Gregory S. Hickok, Department of Cognitive Sciences, University of California-Irvine.

June 2010 – June 2013

Graduate Research Fellow for the National Sciences Foundation

August 2006 – August 2008

Undergraduate Student Researcher for Dr. Frances A. Conners, Department of Psychology, University of Alabama

January 2006 -- December 2007

McNair Scholar Researcher, Dr. Nancy Campbell, McNair Scholars Program, The University of Alabama

Presentations/Published Abstracts

Maddox, C.D., Kislyuk, D., Venezia, J.H., Hickok, G.S. Phonemic and post-phonemic processing of intelligible speech in the anterior and posterior superior temporal sulcus. Society for the Neurobiology of Language Annual Meeting, Amsterdam, Netherlands, August 2014.

Maddox, C.D., Venezia, J.H., Hickok, G.S. Effect of sentence structure on the neural response of auditory cortex. Presentation at the Society for the Neurobiology of Language Annual Meeting, San Sebastian, Spain, October 2012.

Maddox, C.D., Okada, K., Venezia, J.H., Rogalsky, C., Hickok, G.S. Bilateral temporal regions involved in monosyllabic word intelligibility. Presentation for the Neurobiology of Language Conference, Annapolis, Maryland. November 2011.

Maddox, C.D., Venezia, J.H., Hickok, G.S. Functionally hierarchical organization of the superior temporal sulcus in speech perception. Presentation for the Neurobiology of Language Conference, San Diego, California. November 2010.

Maddox, C.D., Venezia, J.H., Hickok, G.S. Functionally hierarchical organization of the superior temporal sulcus in speech perception. Presentation for the Cognitive Neuroscience Society annual meeting, Montreal, Quebec, Canada. April 2010.

Venezia, J.H., Rong, F., Maddox, C.D., Saberi, K., Hickok, G.S. Integration of audiovisual speech does not rely on Broca's area. Presentation for the Cognitive Neuroscience Society annual meeting, Montreal, Quebec, Canada. April 2010.

Loveall, S., Moore, M., Maddox, C.D., Hume, L., Conners, F. An individual difference analysis of the self-teaching hypothesis. Presentation at the Society for the Scientific Study of Reading Annual Meeting, Berlin, Germany, July 2010.

Moore, M. Maddox, C.D., Loveall, S., Hume, L., Conners, F. Possible causes of phonological decoding difficulty in adolescents with intellectual disability. Presentation for the Gatlinburg Conference, Annapolis, Maryland, March 2010.

Maddox, C.D. & Conners, F.A. Reading and intellectual disabilities. Presentation for the Gatlinburg Conference, Annapolis, Maryland, March 2010.

Conners, F.A., Moore, M., Maddox, C.D., Loveall, S., Hume, L. Strengths and weaknesses in reading skills of children with intellectual disabilities. Presentation at the Society for the Scientific Study of Reading Annual Meeting, Boston, Massachusetts, June 2009.

Maddox, C.D. & Conners, F.A. Rhythmic awareness in reading development: The influence of prosodic sensitivity on word identification. McNair Scholars Research Conference, April 2008.

Maddox, C.D. & Conners, F.A. Rhythmic awareness in reading development: The influence of prosodic sensitivity on word identification. Undergraduate Research and Creative Activities Conference, University of Alabama, March 2008.

Maddox, C.D. & Conners, F.A. Rhythmic awareness in reading development: The influence of prosodic sensitivity on word identification. McNair Research Annual Conference, August 2007.

Publications

Loveall, S.J., Maddox, C.D., Conners, F.A. Loveall, S. J., & Conners, F. A. (in press). Individuals with intellectual disability can self-teach in reading. *American Journal on Intellectual and Developmental Disability*.

Conners F.A., Loveall S.J., Moore M.S., Hume L.E., & Maddox C.D. (2011). An Individual Difference Analysis of the Self-Teaching Hypothesis. *Journal of Child Experimental Psychology*, 108(2):402-10.

Maddox, C.D. & Conners, F.A. (2008). Rhythmic Awareness in Reading Development: The Influence of Prosodic Sensitivity on Word Identification. *McNair Scholars Journal*, 8: 103-124.

ABSTRACT OF THE DISSERTATION

Levels of Processing in the Pathways for Intelligible Speech

By

Christopher Dale Maddox

Doctor of Philosophy in Cognitive Neuroscience

University of California, Irvine 2015

Professor Gregory Hickok, Chair

There are divergent claims concerning the broad cortical organization of speechrecognition. One model holds that speech perception and comprehension is governed by a left lateralized anterior temporal lobe (ATL) pathway. Another model argues that bilateral superior temporal regions are critically important, and, in fact, represent a lower level of processing that drives ATL activation in a bottom up fashion. These models were tested in a series of auditory fMRI experiments that gradually investigated lower levels of speech analysis. The experiments contrasted listening to clear monosyllabic words, pseudowords, sentences, and word lists with unintelligible spectrally rotated and time-reversed speech. In the first experiment, posterior temporal regions did not respond differentially to sentence versus word list stimuli, consistent with the idea that bilateral regions of the superior temporal plane support speech recognition at a lower (perhaps phonological) level. An area of the ATL centered around the superior temporal sulcus (STS) was activated more for sentences than word lists, indicating that the region may be involved in sentence-level operations. In the second experiment, this same region in the left hemisphere was activated more by monosyllabic words than rotated words. This suggests that the anterior focus is not exclusively attributable to sentence-level operations. In the third experiment, lexical status was found to differentially modulate anterior and posterior STS regions. There was more activation in the aSTS bilaterally for words than pseudowords, but these conditions did not lead to activation differences in the posterior region. It appears that anterior temporal speech-selective regions respond to lexical-semantic aspects of speech, whereas posterior temporal speech-selective areas are coding lower level phonemic information.

Chapter 1: Introduction

Converging evidence suggests a parallel between the primate and human auditory processing organization. Functional and anatomical studies of the primate auditory system have led to the proposition that the processing of sounds involves separate hierarchically organized processing streams. The posterior pathway is thought to be sensitive to auditory spatial information in primates, while the anterior pathway is sensitive to conspecific communicative sounds (Tian et al., 2001). A similar anterior pathway in humans has been contemplated for the processing of speech sounds, but the precise neural correlates and the degree of lateralization for these processes is of considerable debate. The current series of experiments provides some insight into the cortical organization of speech recognition.

There are divergent claims concerning the broad cortical organization of speech recognition. One model holds that speech perception and comprehension is governed by a left lateralized anterior temporal lobe (ATL) pathway (Rauschecker & Scott, 2009; Scott, Blank, Rosen, & Wise, 2000). This claim is based predominantly on functional imaging studies that manipulate both the acoustic features and the intelligibility of connected speech. In these studies intelligibility was degraded by spectrally rotating the speech signal. The main finding from this line of work is that intelligible speech compared to unintelligible speech produces activation with a peak in the left anterior temporal lobe (Narain et al., 2003; Scott et al., 2000). Another model, derived from functional imaging, neuropsychology, and direct cortical stimulation argues that bilateral *posterior* superior temporal regions are critically important and, in fact, represent a lower (perhaps phonological) level of processing that may then drive higher-level systems in the anterior temporal lobe (Hickok & Poeppel, 2000, 2004, 2007). Evidence for posterior superior temporal involvement in speech perception includes the following. The perception of auditorily presented meaningless syllables compared to various acoustic control stimuli tends to activate the bilateral posterior superior temporal sulcus (pSTS) (Dehaene-Lambertz et al., 2005; Jancke, Wustenberg, Scheich, & Heinze, 2002; Joanisse & Gati, 2003; Price, 2010; Rimol, Specht, Weis, Savoy, & Hugdahl, 2005). Manipulating information at the phonological level of processing in word stimuli also modulates activity in the bilateral pSTS (Okada & Hickok, 2006; Vaden, Muftuler, & Hickok, 2010). And damage to posterior temporal cortex, particularly bilaterally, results in the most severe receptive speech deficits (Bates et al., 2003; Poeppel, 2001).

Many neuroimaging studies have focused on identifying cortical regions that preferentially respond to intelligible speech compared to various acoustically matched controls. These experiments have differed in the types of intelligible stimuli used (sentences, words, phonemes), the types of controls used (rotated, noise-vocoded, time-reversed, foreign language speech, etc.), and in the fMRI protocols (continuous vs. sparse sampling, active vs. passive task). Several neuroimaging studies contrasting monosyllabic words with simple tones or amplitude modulated noise have found significant increases in signal changes along the length of the STS bilaterally with more robust activation in the left hemisphere (Liebenthal, Binder, Piorski, & Remez, 2003; Binder et al., 1996, 1997, 2004; Jancke et al., 2002; Demonet et al., 1992). Similar activation changes have been found when using words and pseudowords compared to time-reversed speech or signal correlated noise (Binder et al., 2000; Price et al, 1996). Further attempts to isolate speech-specific brain mechanisms have used spectrally-inverted speech or unfamiliar (non-native) speech to more appropriately match acoustic features (Obleser et al., 2006, 2007; Liebenthal et al., 2003; Jacquemot et al., 2003; Meyer et al., 2005; Mo¨tto¨nen et al., 2006; Deheane-Lambertz et al., 2005; Desai et al., 2008). Early, and seminal, studies utilizing this approach found that the left aSTS was more responsive to intelligible (vs. rotated) sentences, leading to the hypothesis of a left-lateralized anterior pathway for the recognition of intelligible speech (Narain et al., 2003; Scott et al., 2000).

One problem with most of the neuroimaging studies of speech intelligibility is that they used relatively small sample sizes. It is therefore unclear whether they had the power to detect activation in the entire network. Indeed, a more recent replication of the early spectral rotation studies used a larger sample size and found activation along the entire length of the STS bilaterally (Okada et al., 2010). Another more recent study used time reversed speech as the unintelligible stimulus and also reported strong posterior temporal activation as well as anterior temporal activation, both left dominant (Leff et al., 2008). A second problem is that most of these studies are confounded with respect to the level of analysis that is driving the activation. This is because stimuli are usually comprised of sentences or phrases and when the speech signal is rendered unintelligible, it affects processes at the phonological, lexical, syntactic, combinatorial and prosodic levels of processing. This is particularly problematic for interpreting the ATL activation for two reasons. One is that the ATL has been consistently implicated in sentence-level combinatorial processing (Dronkers, Wilkins, Van Valin, Redfern, & Jaeger, 2004; Friederici, Meyer, & von Cramon, 2000; Humphries, Binder, Medler, & Liebenthal, 2006; Humphries, Love, Swinney, & Hickok, 2005; Humphries, Willard, Buchsbaum, & Hickok, 2001; Rogalsky & Hickok, 2009; Vandenberghe, Nobre, & Price, 2002) as well as higher-order conceptual semantic representation (Patterson, Nestor, & Rogers, 2007; Visser, Jefferies, & Lambon Ralph, 2010). The other, and possibly related, problem is that spectrally rotating the signal tends to have more of an effect on the higher level aspects of comprehension than on the phonological form (Liebenthal, Binder, Spitzer, Possing, & Medler, 2005). Thus, it is possible that phonological processing is carried out in the posterior half of the STS/STG while the anterior activation reflects primarily sentence- or some other non-phonological level process (Hickok & Poeppel, 2007).

The goal of the present studies is to evaluate this possibility by looking within subjects at functional activation to words, nonwords, and sentences in clear, spectrally rotated, and time reversed conditions. Cortical regions that are involved in the perception of phonological information should show similar activation levels for all conditions with phonetic information relative to their rotated counterparts, whereas regions involved in non-phonological processing levels should exhibit different levels of activation for sentences compared to the words and sublexical conditions.

Chapter 2: Experiment 1

The aim of the first experiment was to identify brain regions involved in the perception of speech at two different levels of analysis. To do this we used fMRI to measure activation during listening to intelligible sentences or word lists contrasted with unintelligible rotated versions of the same stimuli. Cortical regions involved in the perception of intelligible speech should show similar activation patterns for both sentences and word lists when compared to their unintelligible counterparts. Regions involved in higher-level, post-lexical processes should have more activation for sentences than word lists.

Materials and Methods

Subjects

Twenty right-handed native English speakers (11 male, 9 female; mean age = 23 years, range 19-31) participated in this study. All participants were free of neurological disease and gave informed consent under a protocol approved by the IRB of UC-Irvine.

Stimuli and Procedures

Participants were presented with 4 different types of auditory stimuli including (1) clear spoken sentences (cS), (2) spectrally-rotated versions of these same sentences (rS), (3) clear spoken lists of words (cL), and, (4) spectrally-rotated versions of these same word lists (rL).

The sentences were derived from items on the BKB sentence list (Bench et al. 1979). Prosodic manipulations in sentences and word lists have been shown to modulate activity in anterior temporal cortex (Humphries et al., 2005). Thus, to control for differences in speech prosody between sentences and lists each word in the set of sentences was recorded individually rather than as connected speech. These words were then concatenated to form structured sentences (corresponding to the BKB sentence set) or were concatenated randomly (to form word lists). The interstimulus interval for both sentences and word lists was 750 msec.

Stimuli were spectrally rotated by multiplying each waveform with a 4kHz tone in the time domain, which results in convolution of their spectra in the frequency domain and, thus, spectral rotation around the 2kHz center frequency axis. The stimuli were low-pass filtered at 3.8kHz before and after the spectral rotation to maintain bandwidth symmetry about the center frequency (Blesser, 1972). This process has been argued to preserve the spectrotemporal complexity of the speech signal (Scott et al., 2000) -- although see Okada et al. (Okada et al., 2010b) -- but renders it unintelligible. All stimuli (157 sentences) were digitally recorded at a sampling rate of 44.1 kHz, and normalized to equal root mean square amplitude.

A single trial was composed of 6 items of a single type (e.g., cS, rS, cL, or rL) that spanned a duration of 12s followed by 1.5 s of silence during which subjects responded. On each trial, participants were asked to indicate with a button press if the speech they heard was intelligible (words were identifiable) or unintelligible. Four trials of each type were randomly presented in each of 8 functional sessions (runs), along with 4 rest trials (scanner noise) per session, for a total of 32 trials per condition in the experiment. A practice session in which each trial (including rest) was presented twice preceded the 8 functional scans, and a high-resolution anatomical scan was obtained last. Data from the practice session were discarded. All stimuli were presented over MR compatible headset and stimulus delivery, and timing was controlled using Cogent software (http://www.vislab.ucl.ac.uk/cogent_2000.php) implemented in Matlab 6 (Mathworks, Inc., Natick, MA).

Scanning Parameters

MR images were obtained in a Philips Achieva 3T (Philips Medical Systems, Andover, MA) fitted with an 8-channel RF receiver head coil, at the Research Imaging Center at the University of California, Irvine. We collected a total of 832 echo planar imaging (EPI) volumes over 8 sessions using Fast Echo EPI (sense reduction factor = 2.4, matrix = 112 x 112 mm, time repetition [TR] = 2.7 s, time echo [TE] = 25 ms, size = 1.957 x 1.957 x 1.5 mm, flip angle = 70, number of slices = 47). After the functional scans, a high-resolution anatomical image was acquired with a magnetization prepared rapid acquisition gradient echo pulse sequence in axial plane (matrix = 240 x 240 mm, TR = 11 ms, TE = 3.54 ms, size = 1 x 1 x 1 mm, flip angle = 18).

Data Analysis

Preprocessing of the data and ROI identification were performed using AFNI software (http://afni.nimh.nih.gov/afni). The anatomical image for each subject was coregistered to the subject's mean EPI image. In each session, the first 3 volumes and the last volume were discarded. Motion correction was performed by creating a mean image from all the volumes in the experiment and then realigning all volumes to that mean image using a 6-parameter rigid-body model (Cox and Jesmanowicz, 1999). Images were then smoothed with an isotropic 6-mm full-width half-maximum (FWHM) Gaussian kernel (smoothed data were used for group analysis only).

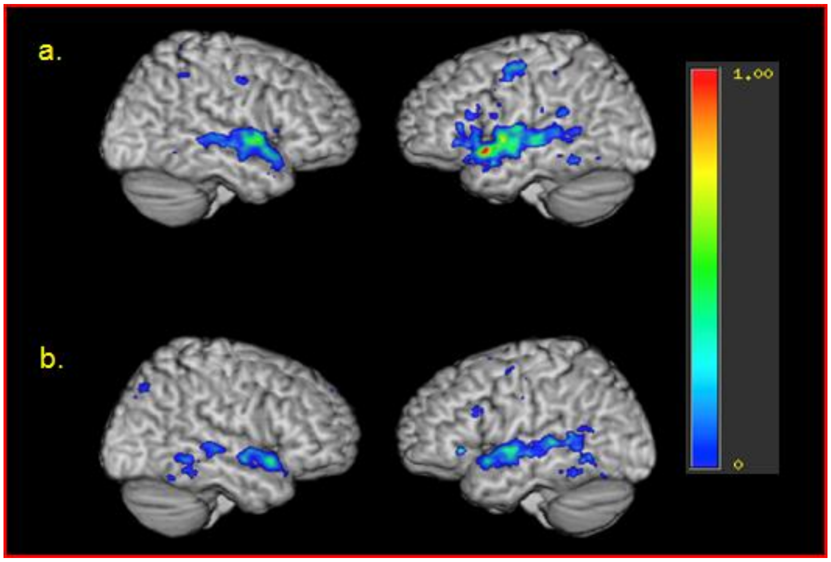
Functional maps for each subject were transformed into standardized space to facilitate group analysis. The images were spatially normalized to a standard EPI template based on the Montreal Neurological Institute (MNI) reference brain (http://www.bic.mni.mcgill.ca/brainweb/) and resampled into 1-mm3 voxels using nonlinear basis functions. Second-level analysis was performed on the linear contrasts of the parameter estimates from each participant, treating participants as a random effect, and voxel-wise t-tests were performed. Statistical threshold in the group analysis was set at $P < 0.001$ FDR.

Results

Subjects were able to identify both clear sentences and word lists as intelligible and spectrally rotated speech as unintelligible with 98% accuracy (s.d. = .035), demonstrating that the stimuli were correctly perceived within the scanner.

Several specific contrasts were performed on the fMRI data. The first sought to replicate previous findings of an anterior focus of activation in the contrast between intelligible sentences and rotated sentences (cS-rS). Consistent with early imaging studies of speech intelligibility, intelligible sentences yielded more activity than unintelligible sentences in left anterior temporal regions. However, consistent with more recent findings, we also noted an increase in the BOLD signal bilaterally and including both anterior and posterior regions of the superior temporal lobe and posterior middle temporal region. Small foci of activation were also found in frontal regions including the IFG and dorsal premotor cortex (Figure 2.1a and Table 2.1). The contrast between intelligible word lists and unintelligible rotated word lists (cL-rL) produced a qualitatively similar pattern of activation, although with a modest posterior shift in the focus of the activation peak for the list stimuli relative to the sentence stimuli (Figure 2.1b and Table 2.1).

Figure 2.1. Group results from a standard BOLD amplitude subtraction analysis (P<.001, FDR) projected onto an MNI template showing regions that respond (a) more to sentences than rotated sentences (cS>rS), and (b) more to lists than rotated lists (cL>rL).



In order to specifically assess the contribution of higher-level syntactic and combinatorial semantic processing on the activation pattern we contrasted the two maps in Figure 2.1 with the subtraction, [cS-rS]-[cL-rL], which isolates regions more activated to sentences (minus their rotated controls) than to lists (minus their rotated controls). This contrast revealed activation foci in the anterior temporal lobes, bilaterally (Figure 2.2) but with little activation difference in the posterior superior temporal sulcus regions that were active in the intelligibility contrasts (cS-rS and cL-rL) and that had previously been implicated in phonological aspects of speech recognition (Hickok & Poeppel, 2007). At the same time, a large swath of anterior temporal regions also did not respond differentially to the sentence versus list conditions relative to their controls suggesting that only some of the anterior temporal activation in previous studies can be attributed to sentence-level processes (Figure 2.3).

Figure 2.2. Group results from a standard BOLD amplitude subtraction analysis (P<.001, FDR corrected) projected onto an MNI template brain showing regions that respond more to intelligible sentences than intelligible lists [cS-rS]-[cL-rL]

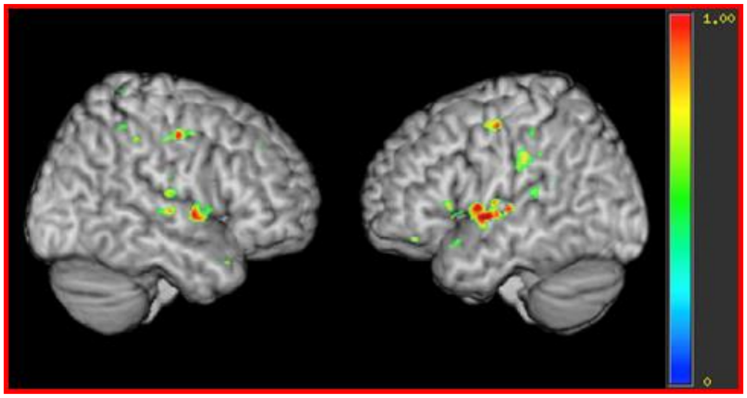
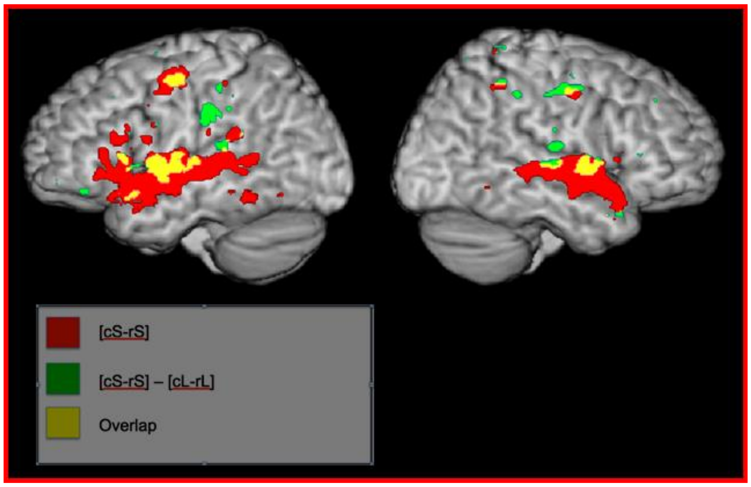


Figure 2.3. Group results from a subtraction analysis (P<.001, FDR) showing regions more active for sentences than rotated sentences [cS-rS] (Red), intelligible sentences than intelligible lists [cS-rS] - [cL-rL] (Green), and both (Yellow).



Discussion

Contrary to the earliest functional imaging studies of intelligibility (Narain et al., 2003; Scott et al., 2000), but consistent with other higher-powered studies (Davis & Johnsrude, 2003; Okada et al., 2010a), we found that intelligible speech activated a bilateral network of superior temporal regions including both anterior and posterior sectors. Sentences and word lists activated a qualitatively similar set of regions. However, when directly compared, sentences activated anterior temporal regions to a greater degree than word lists, demonstrating that a portion of the previously reported intelligibility effect in anterior temporal cortex is attributable to sentence-level language processing features. Posterior temporal regions, particularly the superior temporal sulcus, did not respond differentially to sentence versus word list stimuli (both relative to their rotated control stimuli). The latter finding is consistent with the claim that the posterior superior temporal sulcus bilaterally support speech recognition at the phonological level, which was held constant across the sentence and word list conditions (Hickok & Poeppel, 2007).

Although sentences tended to activate anterior temporal regions more than word lists, we also found that some ATL regions showed an intelligibility effect of equal magnitude for sentences and word lists. This raises the possibility that ATL regions also support lexical and/or phonological level processes (Leff et al., 2008), in contrast to previous claims that ATL regions supported primarily combinatorial processes (Hickok & Poeppel, 2007). The question then becomes, are anterior and posterior regions computationally indistinguishable or are they performing different functions at the lexical or phonological level? The present study cannot answer this question, but there are a number of clues from other work that suggest functional differentiation between posterior and anterior regions. One clue comes from experiments that target sublexical phonological processes in particular and which report primarily posterior STS activation foci (see (Hickok & Poeppel, 2007) for review). An important additional clue comes from the study by Leff and colleagues who used fMRI to identify both posterior and anterior foci in an intelligibility contrast and then used dynamic causal modeling to understand the relation between these nodes in the network. They concluded from their analysis that "auditory speech inputs drive activity in the pSTS; in turn, this activity influences activity in the more anterior areas…" (p. 13214) (Leff et al., 2008). Okada et al. also reported significant differences in the response properties, their sensitivity to acoustic variation in particular, between anterior (more sensitive) and posterior (less sensitive) regions defined by an intelligibility contrast (Okada et al., 2010a). Together, these findings point to some functional differentiation between anterior and posterior regions. While the weight of the evidence comes down in favor of the view that some aspect of phonological-level processing is supported by posterior STS regions, it remains unclear what the anterior regions contribute computationally and at what level, e.g., sublexical, lexical-phonological, lexical-semantic, or something else.

One potential complication with the above discussion is the implicit assumption that listening to concatenated lists of words involves only phonological and lexical processes and does not engage higher-level combinatorial processes. This assumption may not hold, especially given that the word lists were generated by randomly combining words from the entire set of sentences, which results in occasional partial sentence fragments. The implication of this possibility is that the contrast between sentences and word lists may over-estimate the extent of regions identified that show no difference. The fact that we do observe robust differences between the stimulus classes, however, indicates that they are treated differently enough to reveal distinct neural response patterns. The effects of combinatorial processes in the word lists is therefore likely minimal.

Table 1. Active clusters for Experiment 1 grouped by contrasts, reported at p < 0.001, FDR. Coordinates given in Talaraich space.

Contrast	**Region**	**Center of Mass**		
		X	**y**	**z**
CS-RS	*LaSTG*	-55	11	-7
	RaSTG	58	3	1
	LpSTS	-47	-36	0
	RpSTS	46	33	-2
	LmSTS	-59	-23	-1
CL-RL	*RaSTG*	56	12	-8
	LaSTG	-58	-3	0
	RpSTS	49	-30	-8
	LaSTG	-45	10	-9
	LpSTS	-54	-34	4
[CS-RS] - [CL-RL]	*LaSTG*	-59	-5	1
	LaSTS	-44	0	5
	RaSTG	57	-3	2
	L precentral	-53	-30	31
	R precentral	53	-15	44

Conclusions

The response to intelligible spoken sentences relative to an unintelligible baseline activates much of the superior temporal lobe bilaterally including both anterior and posterior regions. This finding in the present study as well as others reviewed above render indefensible previous claims for an anterior left dominant pathway for intelligible speech based on a purely left localized focus of activity (Rauschecker & Scott, 2009; Scott et al., 2000). What remains at issue is the relative contributions of the various subregions within this network. The present study found that only a portion of the anterior temporal activation can be attributed to sentence-level computations. Additional work is needed to parcel out the contributions of phonological, lexical, morphological, semantic, and intonational factors, all of which could be substantial.

Chapter 3: Experiment 2

The second experiment addresses three potential limitations of the first experiment by using individual words presented in isolation instead of sentences or lists in a passive task and presenting the stimuli in silence. Using words in isolation instead of sentences or word lists avoids the problem of combinatorial processes. The perception of sentence stimuli involves combinatorial semantic processes, as well as syntactic and prosodic processes, that are not engaged during the perception of a word in isolation. Furthermore, the second experiment addressed two of the criticisms concerning the Okada et al. (2010) study by using sparse sampling rather than continuous sampling, which avoids acoustic interference from MRI scanner noise, and passive listening rather than an active task (Peelle, Johnsrude, & Davis, 2010).

The aim of the experiment is to expand upon the findings in the first experiment by identifying brain regions that are sensitive to intelligible speech in the absence of robust sentence or phrase level information. The first experiment indicated that a large portion of the temporal lobe in both hemispheres (posterior, anterior, and in between) is sensitive to intelligible sentences (See Figure 2.1). A similar, although less extensive, pattern was observed when comparing intelligible lists of words to their rotated controls.   In the present experiment, individually presented words eliminate the possibility of combinatorial factors, and should reveal cortical regions involved in lexical-semantic processes. The pattern of cortical activation for individually presented words contrasted with rotated words is expected to be even less extensive than what was observed for sentences and lists in Experiment 1, indicating a lower level of speech recognition processes.

Materials and Methods

Subjects

Nineteen right-handed native English speakers (11 male, 8 female; mean age = 24.2 years, range 20-31) participated in this study. All participants were free of neurological disease and gave informed consent under a protocol approved by the IRB of UC-Irvine.

Stimuli and Procedures

Participants were presented with two different types of auditory stimuli: clear spoken monosyllabic words and spectrally rotated versions of those same words.

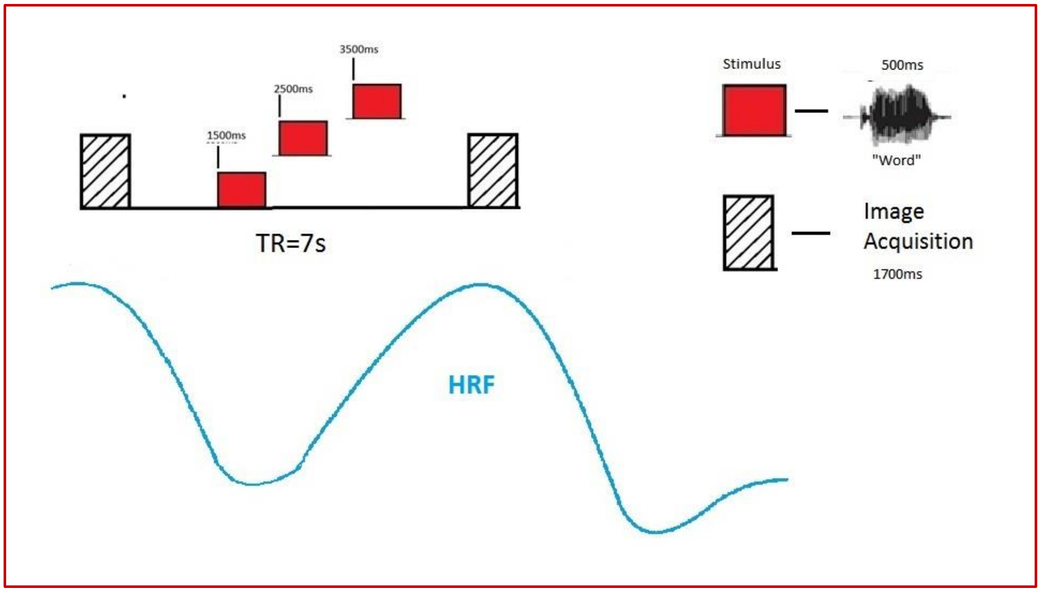
200 words were chosen from IPhOD (http://www.iphod.com.), a database of words and pseudowords developed for speech research that contains phonotactic probabilities as well as density and frequency estimates (Vaden, Halpin, & Hickok, 2009). 50 items were selected from each lexicality, density and phonotactic group. Each group consisted of equal proportions CVCC items and CCVC items, which were chosen for their similar distribution among wide ranges of density and phonotactic values. The mean Kucera-Francis written word frequency for words was 5.29. The mean phonotactic value for low phonotactic items was .0015 and .0034 for high, regardless of lexical or density group. The mean density value was 6.22 for low density words and 14.43 for high.

Recordings were made after pronouncing each stimulus several times to ensure natural speaking rate and clear pronunciation. To optimize the clarity of the speech signal, we used an anechoic chamber equipped with a Shure amplifier and PC running Audacity recording and editing software. Later, Audacity software was used to edit the clearest recording of each item. An analysis of variance detected no significant relationships between conditions and stimuli durations: lexicality, $F(1,392)=.882$, $p=.35$, density, $F(1,392)=.01$, $p=.92$, phonotactic, $F(1,392)=.07$, $p=.79$, and the interactions, F's<.74.

Stimuli were spectrally rotated by multiplying each waveform with a 4kHz tone in the time domain, which results in convolution of their spectra in the frequency domain and, thus, spectral rotation around the 2kHz center frequency axis. As in Experiment 1, the stimuli were low-pass filtered at 3.8kHz before and after the spectral rotation to maintain bandwidth symmetry about the center frequency (Blesser, 1972). Again, this process has been argued to preserve the spectrotemporal complexity of the speech signal (Scott et al., 2000) -- although see Okada et al. (Okada et al., 2010b) -- but renders it unintelligible. All stimuli (157 sentences) were digitally recorded at a sampling rate of 44.1 kHz, and normalized to equal root mean square amplitude.

An event-related design and sparse sampling fMRI acquisition was used to aurally present one word per trial, in the absence of scanner noise. Stimulus presentation was jittered within a 3500ms block so that multiple points of the BOLD response could be sampled (Figure 3.1). Each participant was presented with 200 words and 200 rotated words randomly intermixed with 80 resting trials across 10 runs. There was no active task in the experiment. Participants were asked to stay alert and pay attention to the speech sounds they heard.

Figure 3.1. Schematic illustrating the experimental design for Experiment 2.



Scanning Parameters

MR images were obtained in a Phillips 3T scanner at the University of California, Irvine. After the functional scans, a high resolutional anatomical image was acquired in the axial plane (matrix 256 x 256, TR—8ms, TE—3.7ms, flip angle 8°, size 1 x 1 x 1mm). 49 EPI volumes were collected in each session, matrix—112mm x 112mm, TR—7s, TE—25ms, size—1.957 x 1.957 x 1.5mm, slice thickness—2.5mm, gap--0.5mm, flip angle-- 90. The data were preprocessed and analyzed using AFNI.

Data Analysis

Preprocessing of the data and ROI identification were performed using AFNI software (http://afni.nimh.nih.gov/afni). The anatomical image for each subject was coregistered to the subject's mean EPI image. Motion correction was performed by creating a mean image from all the volumes in the experiment and then realigning all volumes to that mean image using a 6-parameter rigid-body model (Cox and Jesmanowicz, 1999). Images were then smoothed with an isotropic 6-mm full-width half-maximum (FWHM) Gaussian kernel (smoothed data were used for group analysis only).

Functional maps for each subject were transformed into standardized space to facilitate group analysis. The images were spatially normalized to a standard EPI template based on the Montreal Neurological Institute (MNI) reference brain (http://www.bic.mni.mcgill.ca/brainweb/) and resampled into 1-mm3 voxels using nonlinear basis functions. Second-level analysis was performed on the linear contrasts of the parameter estimates from each participant, treating participants as a random effect, and voxel-wise t-tests were performed. Statistical threshold in the group analysis was set at $P < 0.001$ FDR.

Results

Both conditions when compared to rest had similar activation patterns (Figure 3.2, Figure 3.3). Each activated most of the lateral superior temporal gyrus (STG), extending from pSTG to aSTG and into the STS. The activation for Words extended further anteriorly and inferiorly (Figure 3.2). The main effect of intelligibility, revealed by the contrast [Words-rotWords], indicated four clusters significantly more active for words than rotated words (Figure 3.4 and Table 3.1). The largest activation differences were found in the left posterior inferior temporal gyrus. Two significant clusters were observed in the left ATL. The larger of the two clusters was in aSTG, and the smaller was in the frontal operculum. The clusters did not separate at slightly lower thresholds. The final cluster of significant activation was in Broca's Area (BA 44). There were no cortical regions activated more by the rotated words than the words.

Figure 3.2. Group results from a standard BOLD amplitude subtraction analysis (P<.001, FDR) projected onto an MNI template showing regions that respond more to intelligible words than baseline (rest).

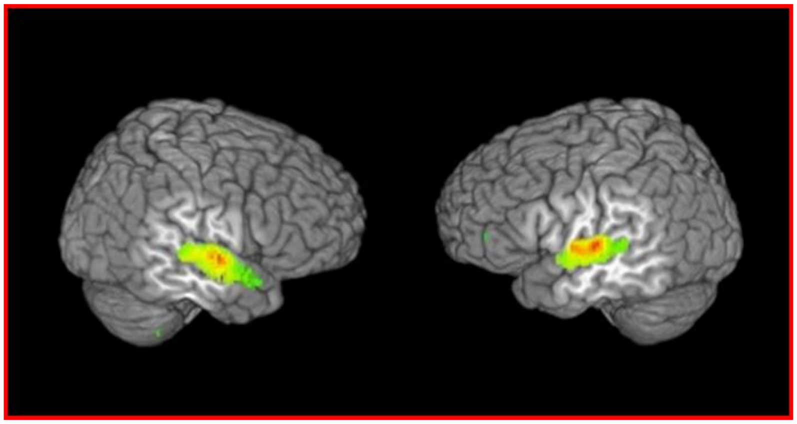


Figure 3..3. Group results from a standard BOLD amplitude subtraction analysis (P<.001, FDR) projected onto an MNI template showing regions that respond more to rotated words than baseline (rest).

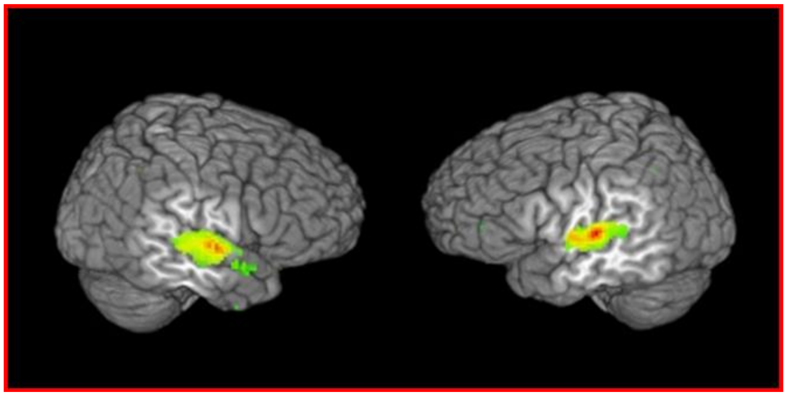


Figure 3.4. Group results from a standard BOLD amplitude subtraction analysis (P<.001, FDR) projected onto an MNI template showing regions that respond more to intelligible words than rotated words (Words>rotWords).

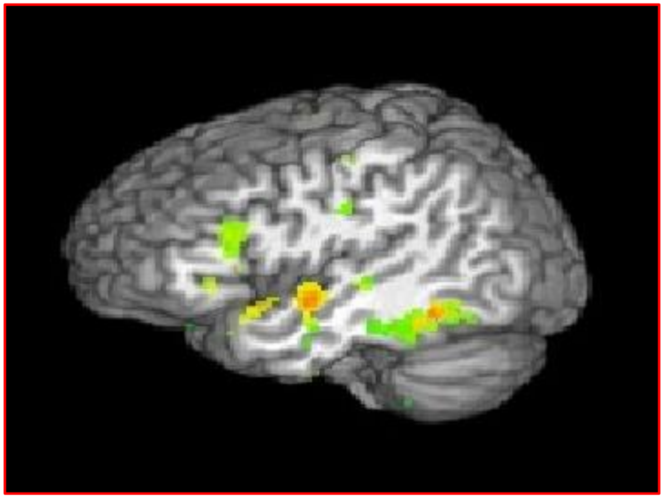


Table 3.1 .Active clusters for Experiment 2 grouped by contrasts, reported at $p < 0.001$, FDR. Coordinates given in Talaraich space.

Contrast	**Region**	**Center of Mass**		
		x	**y**	**z**
W-rotW	*LaSTG*	-49	2	-6
	Temporal pole	-49	12	-11
	IFG/BA 44	-45	14	-12
	pITG	-55	-38	0

Discussion

In contrast with our predictions, the results of this experiment demonstrate a left-lateralized preference for intelligible words. Thus, the present findings provide partial support for the notion of a left-dominant anterior pathway for the processing of intelligible speech. The center of the left aSTG region that was more responsive to intelligible words in this study is very close to the activation focus reported in Scott et al. (2000). This finding indicates that the previously reported activation focus is not exclusively attributable to sentence-level operations. Rather the region appears to be supporting some process that holds at the lexical level or below. The present study cannot pinpoint the source of the effect but possible candidate operations include coding for phonological word forms, semantic processes, and suprasegmental acoustic processes involved in word-level intonational analysis. Each of these levels of analysis differs between words and rotated words.

Although firm conclusions from the present study are under-constrained, the hypothesis that the left aSTS/STG is involved in the initial stages of speech recognition still seems dubious. Evidence from aphasic patients with LaSTS/STG lesions and full ATL lobectomies indicates that anterior regions cannot be critical for phonological word form or semantic operations, because damage (or removal) of the site has little to no effect on word comprehension (Saykin et al., 1995; Bi et al., 2011; Damasio et al, 2004). Patients with left ATL damage perform normally, typically at or near ceiling, on word-picture matching, lexical discrimination, and other semantic comprehension tasks. The linguistic difficulties most commonly associated with the anterior LH damage are in overt naming tasks that require lexical retrieval. These processes are more taxing on memory than cognitive operations related to speech recognition and comprehension. Instead, the neuropsychological evidence suggests that the left pITG (activated in this study) and adjacent regions are more critically involved in single word comprehension (Hart & Gordon, 1990; de Renzi, Zambolin, & Crisi, 1987; Sartori & Job, 1988; Rapcsak & Rubens, 1994). This notion is also supported by various neuroimaging studies that manipulate lexical-semantic attributes of speech stimuli or task-specific cognitive demands (Binder, 2008; Rodd, 2005; Rodd, 2015). LpITG activation in these studies is associated with semantic ambiguity, lexical status (e.g. Word > Pseudoword), conceptual knowledge, and word discrimination. Importantly, the activation changes in the LpITG are consistently related to lexical-semantic operations and unrelated to other linguistic operations commonly attributed to the ATL, e.g. syntactic processes (Rodd, 2015).

The fact that activation differences between the two conditions in this experiment occurred exclusively in the left hemisphere is somewhat surprising, although left-*dominant* activation patterns might be expected. It is possible that the current experimental design, which measured responses to individual stimuli by acquiring only one image per trial, reduced power sufficiently to preclude detection of the broader network activation. Further, the use of a passive task may have caused attentional lapses in our participants. These problems could be avoided by presenting stimuli more frequently, collecting a cluster of images after each trial and using a task that keeps subjects focused. This is what was done in Experiment 3, which also looked more closely at sublexical processes that may be driving temporal lobe activations to speech.

Chapter 4: Experiment 3

A potential limitation of the first two experiments is that all of the intelligible speech stimuli included real words (lexical items), making it difficult to tease apart activations driven by sublexical versus lexical processes, such as lexical access or semantic retrieval. The third experiment expands upon the previous two, as it seeks to further explore levels of analysis in the processing of intelligible speech by including two conditions that tap into processes between low-level acoustic/phonetic analysis and higher level lexical/semantic analyses. The previous experiments identified relatively speech-selective regions in the temporal lobe, revealed by the contrasts between intelligible and rotated speech. The current experiment tests whether these speech-selective regions are sensitive to lexico-semantic attributes of the speech stimuli by comparing words and pseudowords.

Because pseudowords share phonotactic patterns with real words, they may partially activate lexical networks. In consideration of this complication, we also included a condition involving time-reversed words, which sound very speech like, have some local phonological content, but do not follow real word phonotactic patterns (Binder et al. 2000; Pellegrino, Ferrange, & Meunier, 2010). Thus, the important difference between pseudowords and time-reversed speech for the purposes of the current line of investigation is that the pseudowords maintain some degree of word form, whereas the time-reversed speech only maintains partially recognizable phonological chunks.

Time-reversed speech has been variously considered an unintelligible (non-speech) control and a measure of prelexical speech perception. Novice listeners can identify phonemes in reversed speech well above chance, while experts can identify many phonemes with a high degree of accuracy (Binder et al. 2000; Pellegrino et al. 2010). Listeners cannot typically identify individual speech sounds in spectrally rotated speech, nor consistently recognize the sounds as speech (Blesser, 1972; Lachs & Pisoni, 2004). By including pseudowords and time-reversed words with normal monosyllabic words and unintelligible rotated counterparts, the current experiment examines a sort of continuum of lexical and prelexical processes. We hypothesized that (i) anterior temporal regions code speech information at a higher, lexical-semantic level and therefore would respond best to real words, substantially less to nonwords, and quite poorly to the unintelligible speech conditions, whereas (ii) posterior temporal regions code speech information at a lower, phonological level and therefore would respond similarly to words and nonwords, less well but still substantially to time-reversed words, which contain fragments of phonological information, and least well to rotated stimuli.

Materials and Methods

Subjects

Sixteen right-handed native English speakers (7 male, 9 female; mean age = 22.7 years, range 19-35) participated in this study. All participants were free of neurological disease and gave informed consent under a protocol approved by the IRB of UC-Irvine.

Stimuli and Procedures

Participants were presented with 5 different types of auditory stimuli including (1) monosyllabic words (W), (2) monosyllabic nonwords (NW), (3) rotated words (rotW), (4) rotated nonwords (rotNW), and (5) time reversed words (TRw).

180 words and pseudowords were chosen in the same way as in Experiment 2. Rotated versions of these same words were created in the same way as in the first two experiments. Time reversed version of the words and pseudowords were created by flipping the audio signal data in Matlab 6 (Mathworks, Inc., Natick, MA), resulting in stimuli identical to backwards playback of the recorded speech waveforms. The reversal of speech produces nonword stimuli that are similar to the original words in terms of physical complexity and acoustic characteristics. Because many phonemes are relatively temporally symmetrical or show approximate mirror reversal of formant structure before and after a vowel, the reversed speech conveys some phonetic information (Binder et al. 2000; Pellegrino et al. 2010).

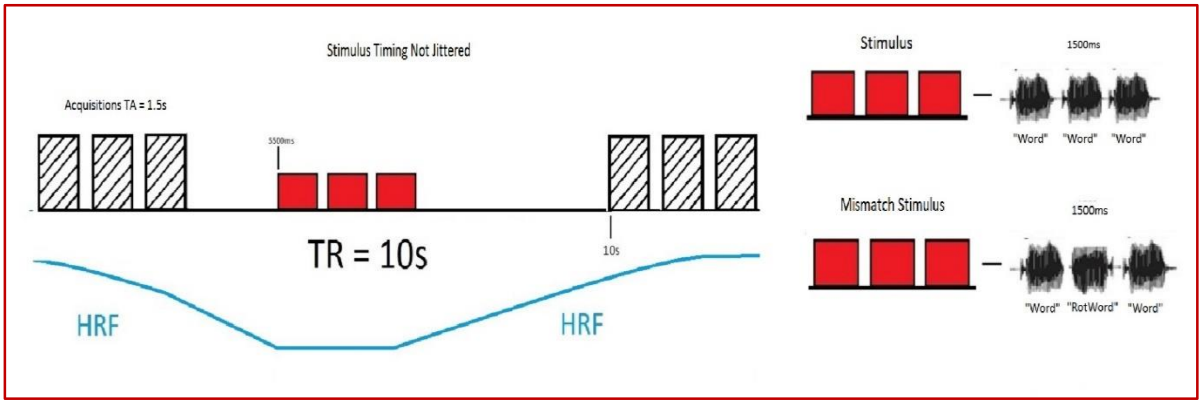
A single trial consisted of 1 item (W, rotW, NW, rotNW, or TRw) repeated 3 times in quick succession (gap = 500 ms) for a total duration of 2.5s. The repetition increases the number and rate of stimulus presentation, which has been shown to maximize effect sizes (Perrachione & Ghosh, 2012). There were also 50 mismatch trials per experiment, in which one of the three items in a triplet was different than the other two. These mismatch trials were task-irrelevant, and were only used to maintain listeners' attentiveness. Participants responded with a button press whenever an item in the triplet differed from its neighbors. However, none of these trials were included in analyses.

An event-related design and clustered sparse sampling fMRI acquisition was used to aurally present the stimuli in the absence of scanner noise. 3 MR images were acquired in rapid succession (TA=1.7s) at the beginning of each 10s TR, 3500ms after the onset of stimulus presentation, and 1000ms after stimulus onset (Figure 4.1). 80 trials per condition plus 50 mismatch trials were randomly presented within 9 functional runs that were 500s in duration (total = 75 min, 1350 EPI volumes). A high-resolution anatomical scan was obtained last. All stimuli were presented over MR compatible headset and stimulus delivery, and timing was controlled using Presentation software (http://www.neurobs.com/presentation/docs/index_html) for Windows (Neurobehavioral Systems, Inc., Berkeley, CA).

Scanning Parameters

Data were collected on the 3T Phillips Achieva MR scanner at the UCI Research Imaging Center. Functional MRI data was collected using single-shot echo-planar imaging (matrix = 112mm x 112mm, TR = 10 s, echo time = 30 ms, TA= 1.5s, flip angle = 70°, voxel size = 2.5 mm × 2.5 mm × 3 mm, SENSE factor = 1.7). MRIcro (Rorden and Brett, 2000) is used to reconstruct the high-resolution structural image, and an in-house Matlab program is used to reconstruct the echo-planar images. After the functional scans, a high-resolution anatomical image is acquired with a magnetization prepared rapid acquisition gradient echo pulse sequence in axial plane (matrix = 240 x 240 mm, TR = 11 ms, TE = 3.54 ms, size = 1 x 1 x 1 mm, flip angle = 18).

Figure 4.1. Schematic illustrating the experimental design for Experiment 3.



Data Analysis

Preprocessing of the data and ROI identification were performed using AFNI software (http://afni.nimh.nih.gov/afni). The anatomical image for each subject was coregistered to the subject's mean EPI image. Motion correction was performed by creating a mean image from all the volumes in the experiment and then realigning all volumes to that mean image using a 6-parameter rigid-body model (Cox and Jesmanowicz, 1999). Images were then smoothed with an isotropic 6-mm full-width half-maximum (FWHM) Gaussian kernel (smoothed data were used for group analysis only).

Functional maps for each subject were transformed into standardized space to facilitate group analysis. The images were spatially normalized to a standard EPI template based on the Montreal Neurological Institute (MNI) reference brain (http://www.bic.mni.mcgill.ca/brainweb/) and resampled into 1-mm3 voxels using nonlinear basis functions. Second-level analysis was performed on the linear contrasts of the parameter estimates from each participant, treating participants as a random effect, and voxel-wise t-tests were performed. Statistical threshold in the group analysis was set at $P < 0.01$ FDR.

ROI Identification and Analysis

In order to address our specific hypotheses about the functional response properties of anterior and posterior superior temporal regions as identified previously, we used data from Experiment 1 to define ROIs that we assessed using the present, independent dataset. Voxels of interest were functionally defined from group-level statistical thresholds. Two ROIs were identified using the peak voxels from intelligibility contrasts in Experiment 1 ([Sent-rotSent] + [List-rotList]). Focal cubic ROIs were drawn around the peak voxels, including 5x5x5 voxels (3 voxels in each direction from peak). The mean Talaraich coordinates of the ROIs were, LaSTS = [-55 11 -7], LpSTS = [-47 -36 0]. Within the ROIs, only voxels with a positive regression coefficient for Sentences compared to rest were included ($P < 0.001$ FDR). A 4x2 two-way ANOVA was performed on the individual subject mean signal values for each condition within both ROIs.

Results

Several specific contrasts were performed on the fMRI data. All conditions when compared to baseline showed significantly more activation along most of the lateral superior temporal gyri (Figure 4.2). The main contrast from Experiment 1 (W-rotW) was repeated and the results were largely replicated with some important additions. Again, the most significant clusters of activation for the contrast were in the left ATL and LpITG (Figure 4.3 and Table 4.1). Unlike the previous experiment, however, the greatest activation (peak) differences were in the left ATL, not the LpITG. Also differing from Experiment 2, significant activation differences were found in the right hemisphere ATL in addition to the left.

Figure 4.2. Group results from a standard BOLD amplitude subtraction analysis (P<.01, FDR corrected) projected onto an MNI template brain showing regions that respond more to all conditions compared to baseline.

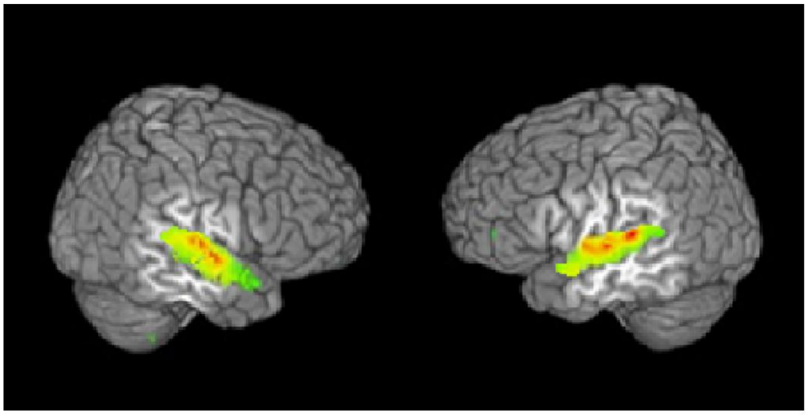
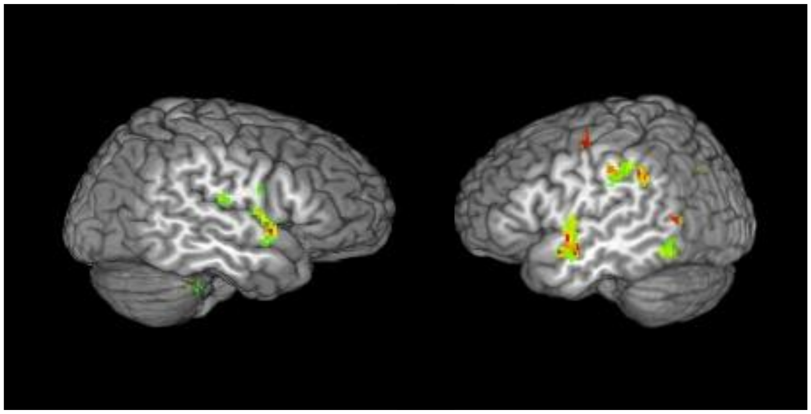
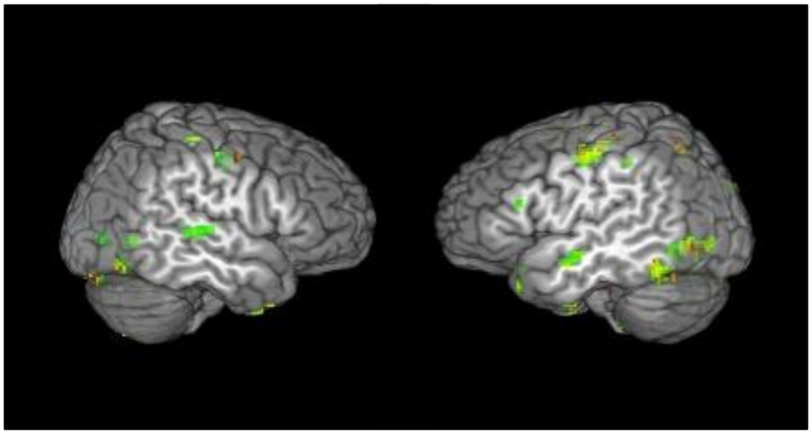


Figure 4.3.. Group results from a standard BOLD amplitude subtraction analysis (P<.01, FDR corrected) projected onto an MNI template brain showing regions that respond more to intelligible words than rotated words (W - rotW).



Experiment 3 had a second (phonologically) “intelligible” condition consisting of non-words (NWs). The left hemisphere aSTS/STG was also activated more by NWs compared to the unintelligible rotated NWs (Figure 4.4). However, this effect was not nearly as robust as the contrast with Words. A region in the right hemisphere STS was also activated more for NWs. This area was not in the ATL, but rather in the middle-posterior STS.

Figure 4.4. Group results from a standard BOLD amplitude subtraction analysis (P<.01, FDR) projected onto an MNI template showing regions that respond more to intelligible non-words than rotated non-words (NW – rotNW).



The direct contrast between the two intelligible conditions (W-NW) revealed robust activation differences in both hemispheres (Figure 4.5). In the left hemisphere, there were two large, yet distinct, clusters in the ATL that were activated more for words. The far anterior region differed from the area activated in the (Word-rotW) and (NW-rotNW) contrasts, while the other (W-NW) cluster partially overlapped these regions. In the RH, the significant activation increases were along the lateral mid-posterior STG and in the pITG.

There were no significant differences in the left hemisphere for the contrast between time-reversed words (TRw) and rotWs (TRw-rotW), indicating that TRw were treated similarly to the other unintelligible speech conditions. However, a portion of the right hemisphere STS/STG was activated more for TRw (Figure 4.6). In the contrast between Words and TRw (W-TRw) the aSTS/STG regions in both hemispheres were again activated more for Words (Figure 4.7).

Figure 4.5. Group results from a standard BOLD amplitude subtraction analysis (P<.01, FDR) projected onto an MNI template showing regions that respond more to intelligible words than intelligible non-words (W – NW).

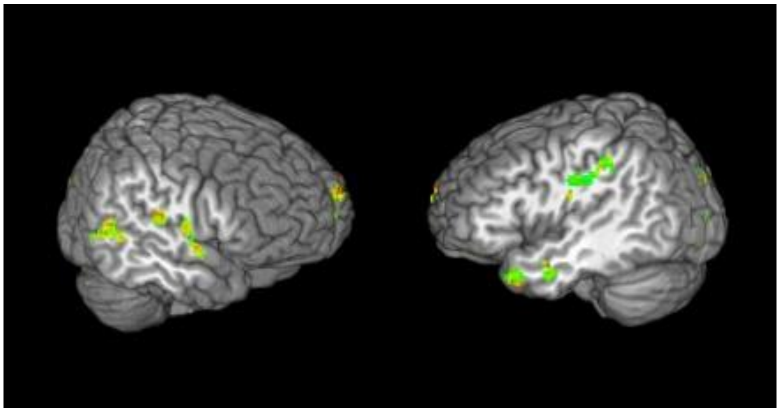


Figure 4.6. Group results from a standard BOLD amplitude subtraction analysis (P<.01, FDR) projected onto an MNI template showing regions that respond more to time-reversed words than unintelligible rotated words (TRw – rotW).

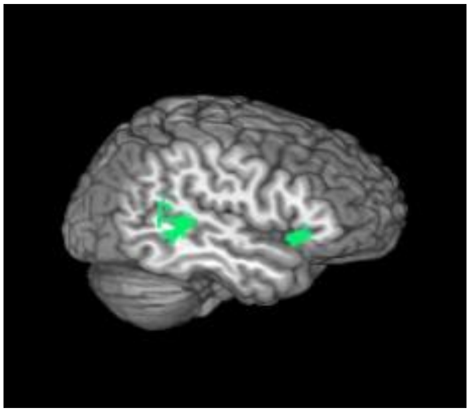
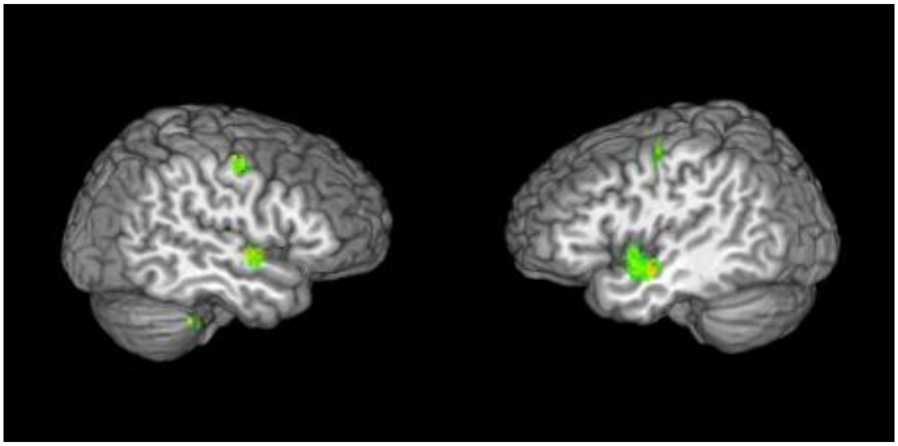


Figure 4.7. Group results from a standard BOLD amplitude subtraction analysis (P<.01, FDR) projected onto an MNI template showing regions that respond more to intelligible words than time-reversed words (W – TRw).



ROI Analysis and Results

ROI analyses was performed on functionally defined voxels from Experiment 1 in the LaSTS and LpSTS. The ROIs only consisted of regions activated more for intelligible speech than unintelligible (rotated) speech. Mean beta-values across voxels in each ROI in each subject were calculated. To estimate the response in each ROI for the “not word” conditions relative to the words, we scaled the beta values in the “not word” conditions to a proportion of the word activation with in each of the two ROIs separately. This allows us to compare relative response to “not word” conditions across ROIs (essentially normalizing within ROI activation to its word response level). We then performed a two-way ANOVA treating Condition (NW, rotNW, rotW, TRw) and ROI (aSTS, pSTS) as factors. The ANOVA revealed a significant main effect of ROI (F(1,120)=54.03, p<0.0001) with LpSTS responding more than LaSTS for each “not-word” condition: NW, rotNW, rotW, and TRw. The main effect of Condition was also significant (F(3, 120)=9.054, p<0.0001), as was the interaction ROI x Condition (F(3, 120)=2.733, p=0.047). Figure 4.8 shows the mean values across each condition. These results partially confirm our hypotheses: aSTS responds significantly less well to nonwords than does pSTS (t(15)=2.13, p=0.02), and minimally to unintelligible speech, whereas pSTS responds strongly to nonwords and with a more graded response drop for unintelligible speech. Contrary to our prediction, however, TRw did not activate pSTS better than rotated conditions.

Figure 4.8. Chart of percent signal change for each condition scaled to the mean (value=1) for Words in each ROI. ROIs were drawn around peak voxels from intelligibility contrast, and only active voxels were included in the ROIs.

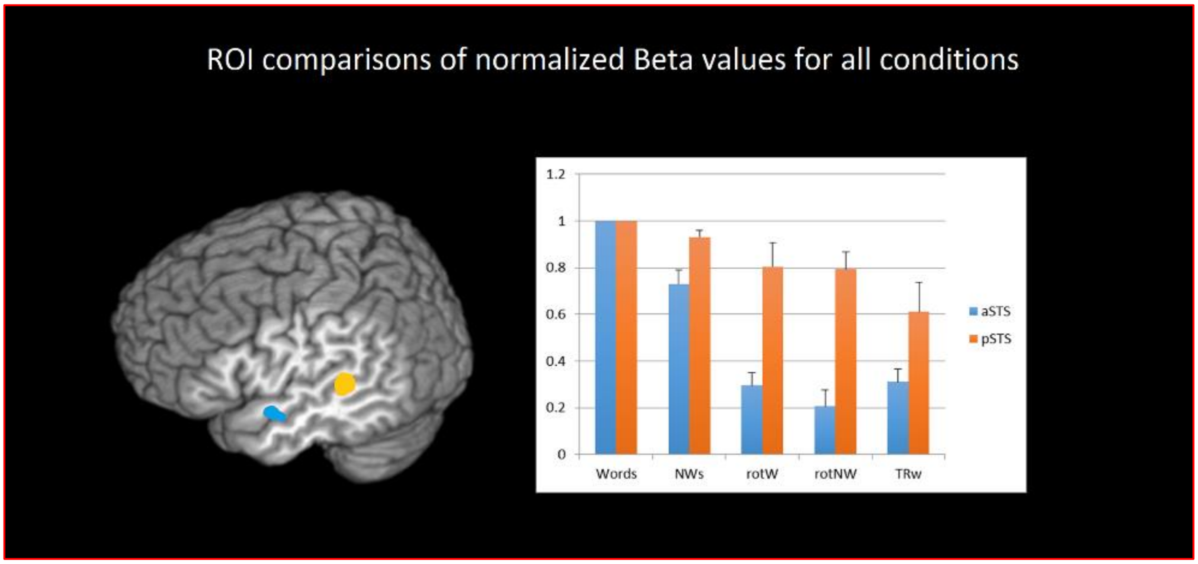


Table 4.1. Active clusters for Experiment 3 grouped by contrasts, reported at p < 0.01, FDR. Coordinates given in Talaraich space.

Contrast	**Region**	**Center of Mass**		
		X	**y**	**z**
W-rotW	*LaSTS*	-57	-5	-3
	Left IPL	-58	-29	20
	LpITG	-51	-42	-10
	RaSTG	59	0	0
NW-rotNW	*LaSTS*	-53	-7	-5
	RmSTG	57	-17	3
	LpITG	-54	-47	-8
	RaSTS	61	-9	1
W-NW	*Temporal pole*	-49	17	-19
	LaSTS	-51	-19	-22
	RpITG	53	-43	-1
	RaSTG	61	-9	1
TRw-rotW	*RpSTG*	58	-32	7
	IFG	55	8	3
W-TRw	*LaSTG*	-53	-4	0
	RaSTG	57	0	-1

Discussion

The previous experiments identified speech-selective areas in anterior and posterior temporal regions. The goal of Experiment 3 was to determine the extent to which these speech-selective areas respond to lexical-semantic and prelexical/phonemic aspects of speech. The results indicate that anterior temporal speech-selective areas respond to lexical-semantic aspects of speech, whereas posterior temporal speech-selective areas are coding lower-level phonemic information. Importantly, lexical status differentially modulated the anterior and posterior regions in the whole brain analysis: the contrast Words-NWs resulted in activation in the anterior region, but these conditions did not lead to activation differences in the posterior region. ROI analysis corroborated these findings showing that the anterior ROI had a larger lexicality effect than the posterior ROI. The increased response to recognizable speech combined with the relative indifference to lexical-semantic content in the pSTS suggests the region is involved in prelexical, possibly phonological processing. This is consistent with many neuroimaging studies of phonological perception (Dehaene-Lambertz et al., 2005; Jancke, Wustenberg, Scheich, & Heinze, 2002; Joanisse & Gati, 2003; Price, 2010; Rimol, Specht, Weis, Savoy, & Hugdahl, 2005) and neuropsychological studies of severe receptive speech deficits (Bates et al., 2003; Poeppel, 2001).

The anterior ROI (LaSTS) had a significantly reduced response (relative to pSTS) for every condition when the two regions were normalized to their respective response to word stimuli. This suggests that activation in the LaSTS area is highly word-selective. The LaSTS also had a significant response to pseudowords (NWs), at least relative to rotated speech. This is presumably because the NWs have word-like form. The words and NWs require much of the same phonological, morphological, and prosodic operations. Therefore, much of the activation in the ATL may reflect common mechanisms. The shared activation may also reflect lexical and conceptual search operations, which ultimately fail when processing pseudowords. It is difficult to tease apart the precise mechanisms common to word and pseudoword processing. However, the results of this experiment indicate that not all of the anterior activity is due to shared mechanisms. At least some of LaSTS response to intelligible speech is due to lexical-semantic operations. It is possible that the aSTS is involved in some type of sound to meaning matching. However, the aSTS does not respond well at all to speech that has been rendered unintelligible, suggesting a reduced role in the spectrotemporal matching associated with the perception of speech sounds.

It is difficult to make any strong conclusions about prelexical and phonological processing with the given results. The best evidence for phonological processing within a certain cortical region would have come from the TRw-rotW contrast, assuming that the participants in the study were recognizing individual phonemes or syllables in the time-reversed words. This contrast did not reveal any significant results in the left hemisphere. The significant activation for TRw in the right hemisphere would be compelling if the same was found bilaterally. The other pre-semantic condition in the experiment consisted of pseudowords (NW). The contrast of NWs with the rotNWs allows for interpretations about speech recognition at a prelexical level. All of the significant activation differences in the contrast were also found for Words, indicating common neural mechanisms and cognitive operations. The usefulness of the NW condition as a reductionist tool with common cognitive subtraction was much greater for the comparisons with Words (W-NW). Significant activation changes for NWs compared to the unintelligible conditions, presumably indicative of lower levels of speech analysis, should, and in fact were also found for Words. Thus, the importance of including NWs in this experiment was in disentangling word-level lexical and semantic processing. The most we can say about sublexical operations is that there are a set of cortical regions that are similarly activated by Words and NWs compared to unintelligible rotated controls. The further conclusion that these regions are involved in the processing of intelligible speech at pre-semantic levels of analysis depends on the W-NW contrast.

The changes in experimental design from Exp.2 to Exp. 3 contributed to substantial changes in the results. The main effect of intelligibility (W-rotW) showed a similar pattern of results in the left hemisphere between experiments. However, the design changes produced significant results in the right hemisphere for Exp. 3, whereas there were no significant RH results for Exp. 2. The design changes also allowed for statistical analysis at higher threshold. The main intelligibility effect patterns were similar at $P < 0.001$ (FDR) for Exp 2 and $P < 0.01$ (FDR) for Exp. 3. It is likely that each alteration in the experimental design contributed to the increases in statistical power to varying degrees, and it is not possible to determine the relative contribution of each. Clustered sparse sampling may be particularly useful for functional imaging of auditory stimuli that are short in duration.

Chapter 5: General Discussion

In the third experiment, we identified anterior temporal regions involved in word-level speech analysis. This finding has important implications for interpretation of the previous experiments' results. In the initial experiment, anterior temporal regions responded more for sentences than word lists, suggesting that the ATL supports post-lexical and sentence-level processing of speech. The second experiment indicated that the left ATL also supports some process that holds at the lexical level. This was demonstrated by comparing the neural response to monosyllabic words in isolation and rotated words. The third experiment further validated that anterior temporal regions support speech recognition at the word-level, and it provided evidence that the regions are sensitive to lexical, and likely conceptual semantic attributes of speech.

Considering the possibility or likelihood that the ATL is involved in word-level lexical and semantic operations, the sentence-level effect in Experiment 1 could also have a semantic component. The sentences and word lists used all of the same lexical items. And, both conditions sounded like lists, in terms of prosody. The primary difference between the conditions concerned contextual meaning, or combinatorial semantics. Thus, it is possible that ATL regions also support sentence-level combinatorial aspects of speech recognition and comprehension. If the ATL regions are critically involved in lexical-level semantic operations, their involvement in higher order semantic processes is intuitively obvious if not necessary. The combined results from the experiments in this study provide compelling support for this notion.

The primary or most troublesome problem with these interpretations is the neuropsychological evidence minimizing the role of the ATL in semantic operations or pointing to other cortical regions as being more critically involved. Two influential theories about the role of the ATL in language processing consider it to be a binding site of the semantic/conceptual properties of words (Patterson, 2007), or that it is crucially involved in lexical retrieval (Damasio, 2004). Several neuropsychological studies have found impaired lexical retrieval with damage or removal of the left ATL, while conceptual knowledge is relatively unaffected (Saykin et al., 1995; Bi et al., 2011; Damasio et al, 2004). Other studies have found difficulties in conceptual processing associated with the left ATL. However, the types of difficulties involve specific details and not general conceptual knowledge (Patterson, 2007). The results of the current studies provide two explanations for intact conceptual processing with left ATL damage. First, we found bilateral ATL activation associated with lexical-semantic operations. So, both hemispheres may contribute to the word-level operations commonly associated with left ATL damage. Secondly, posterior activation of the left inferior temporal gyrus was consistently attributable to lexical-semantic processes in addition to the ATL regions. Thus, both temporal regions may be involved in similar word-level operations.

Both the left ATL and pITG networks are often thought to be involved in the storage and retrieval of conceptual knowledge related to word meaning. Co-activation of the regions in the current study could indicate that the regions are performing similar operations. Or, it can reflect a strong interaction between the two areas in lexical-semantic processing. It is also possible that the ATL and pITG are involved in distinct lexical-semantic operations. Determining the exact nature of word-level operations in the two regions is beyond the scope of the current studies. However, there is evidence of differential processing in the regions. The ATL is implicated more for naming and retrieval tasks, and performance indicates selective difficulties for proper names (Bi et al., 2009). Similar selective difficulties are not associated with LpITG in word comprehension tasks. One possibility is that the ATL is more critically involved in naming and retrieval, and the pITG is more involved in conceptual semantics. Another possibility is that both the ATL and pITG are heavily involved in conceptual (and lexical-semantic) storage and retrieval, but the type of conceptual information processed differs between the regions. The ATL may be involved in more detailed and specific representations (e.g. proper names), whereas the pITG is involved in more general, likely visually influenced and oriented, representations. In this view, neural activity in both regions reflects conceptual search and integrative operations, and not acoustic-phonetic mapping.

Another potential role for the ATL in supporting speech recognition and comprehension is in the transient storage of meaningful acoustic and linguistic information (Peelle & Davis, 2012). The anterior region may facilitate the integration of various segmental and suprasegmental information in speech. It may also allow for predictions to be made for upcoming linguistic content. In this view, the ATL is not critically important for the perception or recognition of speech sounds. Rather, it links memory stores with acoustic and linguistic mapping as sounds are matched to known categories. The anterior activation observed for the pseudowords (NWs) in Experiment 3 (See Figure 4.8) is compatible with this view. The activation may reflect the transient storage of meaningful phonological and morphological units during lexical search. The greater activation observed for Words may reflect a successful search and a need for continued storage.

Further work is need to determine precise cognitive operations associated with neural activity in the left ATL, but the experiments in the current study provide insight into its role in language processing. Most importantly, our results indicate that, (1) the ATL is not involved in acoustic-phonetic mapping, which is a requisite for the perception and comprehension of intelligible speech, (2) word-level and post-lexical processing of intelligible speech is associated with bilateral activation changes in the ATL, and (3) the effects in ATL regions are not exclusively attributable to sentence-level operations; rather, they hold at the lexical level.

REFERENCES

Bates, E., Wilson, S. M., Saygin, A. P., Dick, F., Sereno, M. I., Knight, R. T., et al. (2003). Voxel-based lesion-symptom mapping. *Nature Neuroscience, 6*(5), 448-450.

Bench J, Kowal A, Bamford J (1979). The BKB (Bamford-Kowal-Bench) sentence lists for partially hearing children. *British Journal of Audiology, 13*, 108–112.

Bi Y., Wei T., Wu C., Han Z., Jiang T., & Caramazza A (2011). The role of the left anterior temporal lobe in language processing revisited: evidence from an individual with ATL resection. *Cortex, 47*, 575-587.

Binder J. R., Rao S. M., Hammeke T. A., Yetkin F. Z., Jesmanowicz A., Bandettini P. A., Wong E. C., Estkowski L. D., Goldstein M. D., Haughton V. M., & Hyde J. S. (1994). Functional magnetic resonance imaging of human auditory cortex. *Annals of Neurology, 35,* 662–672.

Binder JR, Frost JA, Hammeke TA, Rao SM, Cox RW (1996) Function of the left planum temporale in auditory and linguistic processing. Brain 119:1239--1247.

Binder JR, Frost JA, Hammecke TA, Cox RW, Rao SM, Prieto T (1997) Human brain language areas identified by functional magnetic resonance imaging. J Neurosci 17:353--362.

Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S., Springer, J. A., Kaufman, J. N., et al. (2000). Human temporal lobe activation by speech and nonspeech sounds. Cerebral Cortex, 10, 512–528.

Binder, J. R., Swanson, S. J., Hammeke, T. A., & Sabsevitz, D. S. (2008). A Comparison of Five FMRI Protocols for Mapping Speech Comprehension Systems. *Epilepsia*, *49*(12), 1980–1997.

Blesser, B. (1972). Speech perception under conditions of spectral transformation. I. Phonetic characteristics. *J Speech Hear Res, 15*(1), 5-41.

Cox, R.W., & Jesmanowicz, A (1999). Real-time 3D image registration for functional MRI. *Magnetic Resonance in Medicine, 42*, 1014-1018.

Damasio H, Tranel D, Grabowski T, Adolphs R, Damasio A (2004). Neural systems behind word and concept retrieval. *Cognition, 92*, 179-229.

Davis, M. H., & Johnsrude, I. S. (2003). Hierarchical processing in spoken language comprehension. *J Neurosci, 23*(8), 3423-3431.

de Renzi, E., Zambolin, A., & Crisi, G. (1987). The pattern of neuropsychological impairment associated with left posterior cerebral artery infarcts. *Brain, 110*, 1099-1116.

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*(1), 21-33.

Demonet, J. F., Chollet, F., Ramsay, S., Cardebat, D., Nespoulous, J. L., Wise, R., et al. (1992). The anatomy of phonological and semantic processing in normal subjects. *Brain, 115*, 1753–1768.

Desai, R., Liebenthal, E., Waldron, E., & Binder, J. (2008). Left posterior temporal regions are sensitive to auditory categorization. Journal of Cognitive Neuroscience 20:7, pp. 1174–1188

Dronkers, N. F., Wilkins, D. P., Van Valin, R. D., Jr., Redfern, B. B., & Jaeger, J. J. (2004). Lesion analysis of the brain areas involved in language comprehension. *Cognition*, *92*(1-2), 145-177.

Freedman, D., & Miller, E. (2007). Neural mechanisms of visual categorization: Insights from neurophysiology. *Neuroscience and Biobehavioral Reviews, 32*, 311-329.

Friederici, A. D., Meyer, M., & von Cramon, D. Y. (2000). Auditory languge comprehension: An event-related fMRI study on the processing of syntactic and lexical information. *Brain and Language, 74,* 289-300.

Grèzes, J. *et al.* (1998) Top-down effect of strategy on the perception of human biological motion: a PET investigation. *Cognit.Neuropsychol.* 15, 553–582.

Hart, J.J., & Gordon, B. (1990). Deliniation of single-word semantic comprehension deficits in aphasia, with anatomical correlation. *Annals of Neurology, 27*, 226–231.

Hickok, G., & Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception. *Trends in Cognitive Sciences*, *4*, 131-138.

Hickok, G., & Poeppel, D. (2004). Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. *Cognition*, *92*, 67-99.

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience, 8*(5), 393-402.

Humphries, C., Willard, K., Buchsbaum, B., & Hickok, G. (2001). Role of anterior temporal cortex in auditory sentence comprehension: An fMRI study. *Neuroreport, 12*, 1749-1752.

Humphries, C., Love, T., Swinney, D., & Hickok, G. (2005). Response of anterior temporal cortex to syntactic and prosodic manipulations during sentence processing. *Human Brain Mapping, 26*, 128-138.

Humphries, C., Binder, J. R., Medler, D. A., & Liebenthal, E. (2006). Syntactic and semantic modulation of neural activity during auditory sentence comprehension. *Journal of Cognitive Neuroscience, 18*(4), 665-679.

Jacquemot C, Pallier C, LeBihan D, Dehaene S, Dupoux E (2003). Phonological grammar shapes the auditory cortex: a functional magnetic resonance imaging study. *Journal of Neuroscience, 23,* 9541-9546.

Jancke, L., Wustenberg, T., Scheich, H., & Heinze, H. J. (2002). Phonetic perception and the temporal cortex. *Neuroimage, 15*(4), 733-746.

Joanisse, M. F., & Gati, J. S. (2003). Overlapping neural regions for processing rapid temporal cues in speech and nonspeech signals. *Neuroimage, 19*(1), 64-79.

Lachs, L., & Pisoni, D. B. (2004). Cross-Modal Source Information and Spoken Word Recognition. *Journal of Experimental Psychology. Human Perception and Performance*, *30*(2), 378–396.

Leff, A. P., Schofield, T. M., Stephan, K. E., Crinion, J. T., Friston, K. J., & Price, C. J. (2008). The cortical dynamics of intelligible speech. *Journal of Neuroscience, 28*(49), 13209-13215.

Liebenthal, E., Binder, J. R., Piorkowski, R. L., & Remez, R. E.(2003). Short-term reorganization of auditory cortex induced by phonetic expectation. Journal of Cognitive Neuroscience, 15, 549–558.

Liebenthal, E., Binder, J. R., Spitzer, S. M., Possing, E. T., & Medler, D. A. (2005). Neural substrates of phonemic perception. *Cerebral Cortex, 15*(10), 1621-1631.

Meyer M, Zaehle T, Gountouna VE, Barron A, Jancke L, Turk A. (2005). Spectro-temporal processing during speech perception involves left posterior auditory cortex. *NeuroReport 16,* 1985–1989.

Mo¨tto¨nen, R., Calvert, G. A., Ja¨a¨skela¨inen, I. P., Matthews,P. M., Thesen, T., Tuomainen, J., et al. (2006). Perceiving identical sounds as speech or non-speech modulates activity in the left posterior superior temporal sulcus. *Neuroimage, 30,* 563–569.

Mummery CJ, Patterson K, Wise RJ, et al. (1999) Disrupted temporal lobe connections in semantic dementia. *Brain, 122*(1), 61–73.

Narain, C., Scott, S. K., Wise, R. J., Rosen, S., Leff, A., Iversen, S. D., et al. (2003). Defining a left-lateralized response specific to intelligible speech using fMRI. *Cerebral Cortex, 13*(12), 1362-1368.

Obleser, J., Boecker, H., Drzezga, A., Haslinger, B., Hennenlotter, A., Roettinger, M., et al. (2006). Vowel sound extraction in anterior superior temporal cortex. *Human Brain Mapping, 27*, 562–571.

Obleser, J., Zimmermann, J., Van Meter, J. & Rauschecker, J.P. (2007). Multiple stages of auditory speech perception reflected in event-related FMRI. *Cerebal Cortex 17*, 2251–2257.

Okada, K., & Hickok, G. (2006). Identification of lexical-phonological networks in the superior temporal sulcus using fMRI. *Neuroreport, 17*, 1293-1296.

Okada, K., Rong, F., Venezia, J., Matchin, W., Hsieh, I. H., Saberi, K., et al. (2010a). Hierarchical organization of human auditory cortex: evidence from acoustic invariance in the response to intelligible speech. *Cerebral Cortex, 20*(10), 2486-2495.

Okada, K., Rong, F., Venezia, J., Matchin, W., Hsieh, I. H., Saberi, K., et al. (2010b). Hierarchical Organization of Human Auditory Cortex: Evidence from Acoustic Invariance in the Response to Intelligible Speech. *Cerebral Cortex, 20*, 2486-2495.

Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience, 8*(12), 976-987.

Peelle JE, Johnsrude IS, & Davis MH (2010). Hierarchical processing for speech in human auditory cortex and beyond. *Frontiers in Human Neuroscience*, 4: 51.

Peelle, J.E., & Davis, M.H. (2012). Neural oscillations carry speech rhythm through to comprehension. *Frontiers in Language Sciences, 3,* 320.

Pellegrino, F., Ferragne, E. & Meunier, F., 2010, "2010, a speech oddity: Phonetic transcription of reversed speech", *Proc. of Interspeech 2010*, 1221-1224.

Perrachione, T., & Ghosh, S. (2013). Optimized design and analysis of sparse-sampling fMRI experiments. *Frontiers in Neuroscience, 7,* 55.

Poeppel, D. (2001). Pure word deafness and the bilateral processing of the speech code. *Cognitive Science, 25*, 679-693.

Price, C.J., Wise, R.J., Warburton, E.A., Moore, C.J., Howard, D., Patterson, K., Frackowiak, R.S., & Friston, K.J. (1996). Hearing and saying. The functional neuroanatomy of auditory word processing. *Brain, 119*, 919-931

Price, C. J. (2010). The anatomy of language: a review of 100 fMRI studies published in 2009. *Ann N Y Acad Sci, 1191*, 62-88.

Rapcsak, S., & Rubens, A. (1994). Localization of lesions in transcortical aphasia. In A Kertesz (Ed.), Localization and neuroimaging in neuropsychology (pp. 297-329). London: Academic Press.

Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nature Neuroscience, 12*(6), 718-724.

Rimol, L. M., Specht, K., Weis, S., Savoy, R., & Hugdahl, K. (2005). Processing of sub-syllabic speech units in the posterior temporal lobe: an fMRI study. *Neuroimage, 26*(4), 1059-1067.

Rodd, J. M., Davis, M. H., & Johnsrude, I. S. (2005). The neural mechanisms of speech comprehension: fMRI studies of semantic ambiguity. *Cerebral Cortex, 15*(8),1261–1269.

Rodd J., Vitello S., Woollams A., & Adank P. (2015). Localising semantic and syntactic processing in spoken and written language comprehension: an activation likelihood estimation meta-analysis. *Brain & Language,* 141, 89-102.

Rogalsky, C., & Hickok, G. (2009). Selective Attention to Semantic and Syntactic Features Modulates Sentence Processing Networks in Anterior Temporal Cortex. *Cereb Cortex*, *19*, 786-796.

Rorden, C., & Brett, M (2000). Stereotaxic display of brain lesions. *Behavioral Neurology, 12***,** 191–200.

Sartori, G., & Job, R. (1988). The oyster with four legs: A neuropsychological study of the interaction of visual and semantic information. *Cognitive Neuropsychology, 5,* 105-132.

Saykin, A. J., Stafiniak, P., Robinson, L. J., Flannery, K. A., Gur, R. C., O’Connor, M. J., & Sperling, M. R. (1995). Language before and after temporal lobectomy: specificity of acute changes and relation to early risk factors. *Epilepsia, 36*, 1071–1077.

Scott, S. K., Blank, C. C., Rosen, S., & Wise, R. J. S. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain, 123,* 2400-2406.

Tian B, Reser D, Durham A, Kustov A, & Rauschecker JP (2001) Functional specializationin rhesus monkey auditory cortex. *Science, 292*(5515), 290–293.

Vaden, K.I., Halpin, H.R., Hickok, G.S. (2009). Irvine Phonotactic Online Dictionary, Version 2.0. [Data file]. Available from http://www.iphod.com.

Vaden, K. I., Jr., Muftuler, L. T., & Hickok, G. (2010). Phonological repetition-suppression in bilateral superior temporal sulci. *Neuroimage, 49*(1), 1018-1023.

Vandenberghe, R., Nobre, A. C., & Price, C. J. (2002). The response of left temporal cortex to sentences. *Journal of Cognitive Neuroscience, 14*(4), 550-560.

Visser, M., Jefferies, E., & Lambon Ralph, M. A. (2010). Semantic processing in the anterior temporal lobes: a meta-analysis of the functional neuroimaging literature. *J Cogn Neurosci, 22*(6), 1083-1094.

Zatorre RJ, Evans AC, Meyer E, Gjedde A (1992) Lateralization of phonetic and pitch discrimination in speech processing. *Science, 256*, 846-849.