NEURAL CORRELATES OF PHONETIC CATEGORY STRUCTURE:

AN FMRI INVESTIGATION

BY

EMILY B. MYERS

B.A., UNIVERSITY OF IOWA, 1999

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY IN THE DEPARTMENT OF COGNITIVE AND LINGUISTIC SCIENCES AT BROWN UNIVERSITY

PROVIDENCE, RHODE ISLAND

MAY, 2006

This dissertation by Emily B. Myers is accepted in its present form
by the Department of Cognitive and Linguistic Sciences as satisfying the
dissertation requirement for the degree of Doctor of Philosophy.

Date________________ ____________________________________
Sheila Blumstein, Advisor

Recommended to the Graduate Council

Date________________ ____________________________________
James Morgan, Reader

Date________________ ____________________________________
Philip Lieberman, Reader

Approved by the Graduate Council

Date________________ ____________________________________
Sheila Bonde, Dean of the Graduate School

Curriculum Vitae

Emily B. Myers

Department of Cognitive and Linguistic Sciences
Brown University
Box 1978
Providence, RI 02912

Office: 137 Metcalf Research
Phone: 401-863-7254
E-mail: Emily_B_Myers@Brown.edu

Education

August, 2005	Ph.D. in Cognitive Science, Brown University (expected) Dissertation title: “Neural correlates of phonetic category structure: An fMRI investigation
May, 1999	B.A. with highest distinction in Spanish and Linguistics, University of Iowa

Honors and Awards

2004-2005	Dana Fellowship
2003-2005	NIH National Research Service Award to Emily Myers, "Neural Bases of Lexical Effects"
1999-2000	Brown University Fellowship
1999	Sanxay Prize for promise of achievement in graduate work
1996-1999	Presidential Scholarship, University of Iowa
1996-1997	State of Iowa Scholar

Publications

Hutchison, E., Blumstein, S.E., and Myers, E.B. (submitted) An event-related fMRI investigation of voice-onset time discrimination.

Prabhakaran, R., Blumstein, S.E., Myers, E.B., and Hutchison, E. (submitted) An event-related fMRI investigation of phonological-lexical competition.

Blumstein, S.E., Myers, E.B, and Rissman, J. (2005). The perception of voice-onset time: An fMRI investigation of phonetic category structure. *Journal of Cognitive Neurosicence* (in press).

Myers, E.B. and Blumstein, S.E. (2005). Selectional restriction and semantic priming effects in normals and Broca's aphasics. *Journal of Neurolinguistics* 18(3), 277-296.

Conference Presentations

Myers, E.B. and Blumstein, S.E. (2005). Neuroimaging evidence for effects of lexical status on phonetic discrimination. *Poster presented at the 12th annual meeting of the Cognitive Neuroscience Society, New York, NY.*

Myers, E.B. and Blumstein, S.E. (2004). The Perception of Voice-Onset Time: An fMRI Investigation of Phonetic Category Structure. *Poster presented at the 11th annual meeting of the Cognitive Neuroscience Society, San Francisco, CA.*

Myers, E.B., Zosh, W.D., Eliassen, J.C., & Sanes, J.N. (2003). Cross-modal modulation of human extrastriate visual cortex by concrete words. *Paper presented at the 33rd Annual Meeting of the Society for Neuroscience, New Orleans, LA.*

Teaching Experience

2003-2005	Sherdian Center Teaching Consultant
Fall 2003	Teaching Assistant, Introduction to Cognitive Science (CG1)
Fall, 2003	Teaching Assistant, Language and the Brain (CG148)
Spring, 2002	Teaching Assistant, Language, Truth and Advertising (CG7)
Fall, 2000	Teaching Assistant, Production, Perception, and Analysis of Speech (CG123)

Acknowledgments

First and foremost, I am grateful to my advisor, Sheila Blumstein, for her guidance, her good humor, and her impressive intellectual curiosity. She has been the best possible role model, mentor, and friend one could hope for, and has been a true inspiration to me. My committee members, Jim Morgan and Phil Lieberman, have offered valuable alternative perspectives on my research, and have both been extremely supportive and helpful.

Learning the subtleties of fMRI design and analysis would never have been possible without the help of Jerome Sanes and Jim Eliassen, who both responded to innumerable questions about design, analysis and interpretation of the studies in this thesis. Emmette Hutchison was always willing to help me debug another script or debate the finer points of AFNI, and graciously gave up many hours of his weekends to help me scan subjects at the magnet, and for that I thank him as well.

I would also like to thank the many members of the Speech Lab I have worked with over the years. Special thanks are due to John Mertus and Kathy Kurowski, who have been enormously helpful as well as good friends and colleagues.

Katherine White, Eiling Yee, and Jesse Hochstadt deserve thanks for always being available for advice and moral support. They and a number of friends, fellow graduate students and officemates have made it fun to come to work every day. Thanks to Marianne Harrison, Wendy Zosh, Quoc Vuong, Ana Quirino Simoes, Jean Vettel, Giulia Righi, Max DiLuca, and Michele Hebert.

Finally, I am grateful to my husband, Paul Allopenna. Without his support, understanding, and encouragement, this dissertation simply would not have been possible.

Table of Contents

List of Tables

List of Figures

1. CHAPTER 1: INTRODUCTION

The process of mapping from an acoustic signal to a meaningful semantic construct is a computationally complex problem. Even early stages of this process, in which acoustic signals are mapped to the phonetic categories that are relevant for a given language, present near-intractable problems for sophisticated machine speech recognition algorithms. It is therefore of interest to determine how humans achieve this mapping so effortlessly, and more specifically, what neural mechanisms underlie this ability. It is the goal of this research to elucidate the neural systems responsible for phonetic categorization. To this end, a series of studies will be reported measuring brain activation levels using fMRI for subjects engaged in phonetic categorization tasks using stimuli which vary along an acoustic-phonetic continuum.

In order to investigate this issue, it is first necessary to elucidate a theoretical framework, and consider the proposed instantiation of this framework with respect to the neural system underlying it. To this end, the cognitive architecture of the speech-lexical processing system and the various processes required to map from acoustics to lexical form will be reviewed. We will then discuss the proposed relationship between this system and its neural basis in the context of current neural models of speech processing. Finally, we will consider methodological issues involved in the use of fMRI. In particular, the pros and cons of this method will be discussed as a backdrop to the fMRI studies presented in this thesis.

1.1. Theoretical framework: Cognitive architecture of the system.

Most current models of language processing assume that the incoming acoustic input must undergo several stages or levels of processing before being comprehended as meaningful speech (McClelland & Elman, 1986; Norris, 1994; Norris, McQueen, & Cutler, 1995). A listener must extract basic acoustic properties of the stimulus, such as fundamental frequency, onset and offset frequency patterns, and amplitude maxima and minima. Formant frequencies may be extracted at this stage, inferred from a speaker's knowledge of, among other things, the possible patterns of formants given the range of human vocal tract sizes, glottal source characteristics, and the fundamental frequency of the speaker. It is hypothesized that this basic level acoustic information is fed to a phonetic category processing stage, in which the basic cues to the identity of the speech sound are extracted. In models such as TRACE (McClelland & Elman, 1986) which allow for competition between possible representations between and within each level of processing, a phonetic category may compete for selection with other similar sound segments at this stage of processing. This information is in turn fed to a lexical-phonological stage where word selection takes placed based on a match between sound structure information and a stored lexical representation. As in the phonetic stage of processing, words with similar phonological structure may compete at this stage of processing. The extent of competition is a function of the similarity of the sound structure and, as well, the number of words which are phonologically similar to each other (Gaskell & Marslen Wilson, 1999; Luce & Pisoni, 1998).

Current models of speech processing differ with respect to the nature of the representation which is passed between levels of processing. Models such as TRACE and the distributed Cohort model (Gaskell & Marslen Wilson, 1997, 1999) allow for graded activation of linguistic units to be passed between levels. For instance, a degraded exemplar of the sound segment /p/ will more weakly activate the /p/ phonetic category unit than will a good exemplar of that category. Likewise, if the degraded /p/ is embedded in a word like “pear”, the word unit “pear” will be more weakly activated than it would be if the initial phoneme were a better example of the sound /p/. In contrast, in other models (c.f. Phillips, 2001) information from the phonetic level of processing is passed to a separate phonological level of processing. At this stage, the phoneme which represents the best match between the input and a phonetic category is chosen. Then this phonological representation is fed forward to the lexical stage of processing. Importantly, any within-category variability inherent in the phonetic representation is discarded at the phonological level of processing. Therefore, such models posit that variability in the acoustics of a phonetic token should have little effect on the eventual access of a word, so long as that variation is variation within a phonetic category. Evidence exists from behavioral paradigms that effects of graded activation of the input can be seen at higher levels of processing (Andruski, Blumstein, & Burton, 1994; McMurray, Tanenhaus, & Aslin, 2002; McMurray, Tanenhaus, Aslin, & Spivey, 2003; Utman, Blumstein, & Sullivan, 2001). Of interest is the extent to which such graded activation may be seen along the neural processing stream.

Models of language comprehension also differ with respect to how much feedback is permitted between levels of processing. In some models (Norris, 1994; Norris et al., 1995), no top-down feedback is permitted from higher levels of processing on to lower levels of processing. In contrast, in interactive language architectures (Dell, Schwartz, Martin, Saffran, & Gagnon, 1997; McClelland & Elman, 1986) feedback is permitted between levels. For instance, lexical information, once available, may be used to bias the processing at the phonetic level of analysis. Ganong (1980) showed presented subjects with stimuli which ranged along an acoustic-phonetic continuum from a word to a non-word (e.g. gift—kift) or a non-word to a word (e.g. giss-kiss). Subjects' performance on a phonetic categorization task showed that the presence of a lexical item at one end of the continuum shifted the phonetic category boundary such that there were more responses which formed a real word. This effect has been attributed to top-down influences of lexical information on the phonetic categorization process. Models which assume a unitary mapping stage from the speech stream to the lexicon can also accommodate lexical effects on perception. In the distributed Cohort model (Gaskell & Marslen Wilson, 1997, 1999), a very basic representation of the speech stream is mapped directly onto a lexical-phonological representation. In such a model, a calculation of the fit between the incoming speech stream and a stored lexical representation should bias the perception of an ambiguous speech sound towards an interpretation of that sound which is consistent with a word in the lexicon.

In the studies reported here, we assume that separate levels of processing exist for acoustic, phonetic-phonological, and lexical processing. Neuroimaging evidence will be brought to bear on two aspects of these models. Experiments 1 and 2 will explore the extent to which graded activation patterns emerge along the neural processing stream. Experiment 3 will examine evidence for feedback between lexical and phonetic-phonological levels of processing.

1.2. Phonetic category structure and voicing in stop consonants

A good deal of acoustic variation is inherent within each phonetic category. As an example, voice onset time (VOT) is an acoustic parameter which contributes to the perception of the voicing distinction in stop consonants (Lisker & Abramson, 1964). VOT is defined as the time between the release of the stop closure and the onset of phonation. While in English initial stop consonants with VOTs of greater than ~30 msec are usually perceived as voiceless, and those with VOTs less than 30 msec are typically heard as voiced stops, the actual VOTs produced by speakers vary considerably within a given voicing category. An important question for speech perception research is how listeners resolve this variation in order to extract a stable phonemic percept. Liberman and colleagues (Liberman, Harris, Hoffman, & Griffith, 1957; Liberman, Harris, Kinney, & Lane, 1961) first observed that listeners are relatively poor at discriminating phonetic exemplars which fall within a phonetic category, yet are very good at discriminating exemplars which cross a category boundary. This phenomenon was labeled the ‘categorical perception’ phenomenon.

Despite this finding, it has been demonstrated that listeners are sensitive to differences within a phonetic category. Using an AX discrimination paradigm, Pisoni and Tash (1974) showed that “same” discrimination responses were slower for acoustically different tokens which fell within a phonetic category than for “same” responses to two identical stimuli. That is, even when subjects identified the stimuli as members of the same category, and were not aware of differences between the stimuli, reaction times were slower when such differences were present. Additionally, explicit judgments of phonetic category “goodness” show structure within a phonetic category. For instance, Kuhl (1991) showed that subjects rated exemplars of the vowel /i/ as poorer exemplars when they were further away in acoustic space from the category prototype. Similarly, Miller and Volaitis (Miller & Volaitis, 1989; Volaitis & Miller, 1992) showed that subjects rated voiceless stops with very long VOTs as poorer exemplars of the voiceless phonetic category.

In the following studies, the nature of phonetic category structure and the neural systems which underlie sensitivity to such structure are examined using a phonetic categorization task to speech tokens which vary along a VOT continuum as participants' brain activation levels are measured by fMRI. The phonetic categorization task is useful for two reasons. One is that it provides a means of isolating the neural processes involved in processing variability in phonetic category structure. That is, it is possible to examine potential differences in activation patterns as a function of the nature of different exemplars of the phonetic category. The second is that the behavioral performance observed in this task typically shows both categorical-type responses to stimuli as well as more graded responses to such stimuli. The identification function yielded when a subject categorizes stops which vary along a voicing continuum typically shows a very sharp boundary between stop categories, which would seem to indicate a finely tuned division between phonetic categories. Nonetheless, reaction times for items within a phonetic category vary as a function of both the "goodness of fit" of that item to its phonetic category as well as the competition between phonetic categories. Thus, it is possible to compare potential differences in activation levels to items along the phonetic continuum that vary either between phonetic category or within phonetic category *for the same task*.

1.3. Neuro-functional architecture of the language processing system

Lesion models of language processing have long highlighted the role of both posterior and anterior brain structures in speech processing. Nearly all aphasics present with some deficit, either receptive or productive, in processing the sounds necessary for language (Blumstein, 1998). These deficits range from dense auditory comprehension disorders, in the case of pure word deafness (Poeppel, 2001), to disorders of phoneme identification and discrimination, in the case of Broca's and Wernicke's aphasia. Of importance, disorders in processing the sounds of speech are not restricted to those patients with severe comprehension deficits, nor does every patient with a severe comprehension disorder necessarily have a phonological processing deficit. For example, Broca's aphasics, who exhibit relatively spared comprehension, typically show abnormal performance on tasks of phoneme identification (e.g. Basso, Casati, & Vignolo, 1977; Blumstein, Cooper, Zurif, & Caramazza, 1977) and, to a lesser extent, phoneme discrimination (e.g. Blumstein, Baker, & Goodglass, 1977; Blumstein, Cooper et al., 1977). Neuroimaging research on the nature of the speech processing system was originally concerned with identifying broad brain regions that underlie language function generally (Demonet et al., 1992; Howard et al., 1992; Petersen, Fox, Posner, Mintun, & Raichle, 1988, 1989; Price et al., 1992). Attention shifted to more specific questions about neural responses to specific stimulus types (e.g. speech versus non-speech), and under specific task demands (e.g. tone discrimination vs. phoneme discrimination) (Belin & Zatorre, 2000; Belin, Zatorre, & Ahad, 2002; Binder et al., 2000; Giraud et al., 2000; Zatorre, Evans, Meyer, & Gjedde, 1992). A wide variety of brain regions, both in the left hemisphere and the right, have been reported as activated for language tasks, including pre-frontal cortex, middle temporal cortex, the supramarginal and angular gyri, the right cerebellum, the thalami, and even Brodmann area 18, which is activated in tasks which involve visual presentation of word stimuli. However, the two regions where activation emerges most consistently in speech-language tasks are the bilateral superior temporal gyri and the left inferior frontal gyrus.

1.3.1. Speech perception and the superior temporal gyri

In current neural models of the speech processing stream, auditory information is relayed via the medial geniculate nucleus to primary auditory cortex (Heschl’s gyrus, also called the transverse temporal gyrus) bilaterally. Here, the signal undergoes basic acoustic processing. This information is then fed to the auditory association cortex which surrounds Heschl’s gyri posteriorly, anteriorly, and laterally in the superior temporal gyrus. Evidence exists from lesion studies on pure word deafness that the superior temporal gyri bilaterally may be the most basic substrate of speech perception (Poeppel, 2001).

Patients with pure word deafness are markedly impaired in comprehension, yet are not truly deaf—unlike patients with cortical deafness, they show intact frequency discrimination and their ability to process non-speech auditory stimuli (such as music) are unimpaired (Poeppel, 2001), although more often patients present with impairments in both speech and non-speech sounds. These patients show normal speech and language production, however, and their impairment seems to be confined to low-level processing of speech stimuli, as shown by intact reading abilities. Unlike auditory agnosics, pure word deaf patients' recognition of familiar sounds is typically intact. They show moderate to severe deficits in consonant identification, yet intact or near intact identification of vowels, which suggests a disruption in the temporal processing of rapidly changing stimuli like the bursts in stop consonants (Auerbach, Allard, Naeser, Alexander, & Albert, 1982; Miceli, 1982; Poeppel, 2001; Saffran, Marin, & Yeni Komshian, 1976). Pure word deafness results from *bilateral* cortical and subcortical lesions of the superior temporal lobes, specifically in posterior regions (H. Damasio, 1998; Poeppel, 2001) or deep unilateral left hemisphere lesions that cut off projections to the left hemisphere (Engelien & Huber, 2001; Geschwind, 1965).

Griffiths and colleagues (Griffiths, Rees, & Green, 1999) propose that auditory agnosia can also be accounted for by the same deficit in complex sound processing as pure word deafness. They propose that pure word deafness is not a distinct pathology from auditory agnosia, and that supposed pure word deaf patients have disorders in identifying not only speech sounds, but also complex non-speech sounds, particularly those which involve fine temporal resolution. This hypothesis is supported by research showing that pure word deaf patients are significantly impaired on rapid temporal transitions in speech and non-speech sounds, yet unimpaired on steady-state stimuli (Wang, Peach, Xu, Schneck, & Manry, 2000). Additionally, most patients with dense comprehension deficits of this sort also exhibited some additional deficit in music processing (Eustache, Lechevalier, Viader, & Lambert, 1990; Takahashi et al., 1992; Tanaka, Yamadori, & Mori, 1987) or both music and environmental sound agnosia (Auerbach et al., 1982; Motomura, Yamadori, Mori, & Tamaru, 1986). Pure word deafness in the absence of any other auditory processing deficit appears to be extremely rare, if not totally non-existent. Conservatively speaking, therefore, the posterior temporal lobes may be said to have a role in processing of complex auditory stimuli, and may or may not contain areas specific to speech processing. Neuroimaging results support this view. Cortical areas in the superior temporal sulcus and on the lateral surface of the STG, ventral to primary auditory areas, appear to specialize in complex auditory stimuli of many kinds, including the spectral shifts characteristic of speech sounds as well as complex non-speech sounds (Belin et al., 2002; Belin, Zatorre, Hoge, Evans, & Pike, 1999; Belin, Zatorre, Lafaille, Ahad, & Pike, 2000; Zatorre & Belin, 2001).

At the level of the superior temporal gyri, models differ with respect to the direction of flow of information, and the nature of speech processing which takes place in the STG. Hickock & Poeppel (2000) propose that the posterior portion of the bilateral STG are responsible for “constructing sound-based representations of speech” (136). From there, they propose, two streams of processing diverge. A left-lateralized dorsal stream proceeds posteriorly and dorsally towards inferior parietal areas, then projects forwards to inferior frontal areas where articulatory representations of speech sounds are referenced. This stream, they propose, is responsible for auditory-motor integration, which they argue is recruited in sub-lexical speech tasks such as phoneme monitoring or phonetic categorization. The ventral stream of processing continues posteriorly and ventrally from the posterior STG to the left posterior middle temporal gyrus and angular gyrus, where they propose there is an interface between sound and meaning. In contrast, Scott & Johnsrude (2003) propose that information flows both anteriorly and posteriorly along the STG from Heschl’s gyri, with anterior portions handling complex spectro-temporal processing of auditory signals generally, and posterior portions contacting articulatory representations of phonetic categories. Like Hickock & Poeppel, they propose that the posterior (dorsal) stream reflects contact with articulatory representations of speech sounds. Unlike Hickock & Poeppel, however, they propose that explicit semantic processing occurs in the anterior stream, which terminates in the left IFG.

In both of these models, parallels are drawn between the two streams of processing and two competing theories of speech perception: feature-based perception versus gestural (motor theory) theories of speech perception. In both models, the posterior/dorsal stream is implicated in contact between the incoming acoustic signal and articulatory representations of this signal, which, both sets of authors suggest, is maximally recruited in tasks which involve explicit attention to phonetic units. In both models, these streams terminate in the left inferior frontal lobes. Given the somewhat uncontroversial view that left inferior frontal areas such as the IFG and insula are involved in speech production abilities (Dronkers, 1996), it is reasonable to suppose a productive role for these areas in the language system, be it in specifically accessing articulatory representations which are proposed to be necessary to explicitly identify a phoneme, or something more incidental to the task, such as subvocal rehearsal.. However, it is unclear from these models whether this “articulatory” stream is utilized at all during natural speech comprehension, or only during tasks which require explicit attention to phonetic structure.

The two models differ widely in the neural substrate for sound-to-meaning mappings, the system which in both views is responsible for most normal receptive speech ability. Hickock & Poeppel argue that this process is mediated by structures posterior and ventral to the STG, with the sound structure of speech contacting lexical-semantic representations in the temporo-parietal-occipital junction. Some evidence exists to suggest that this area may have a role in mediating lexical representations. Wernicke's aphasics, who typically have damage that extends posteriorly from the left STG, typically present with some comprehension impairment on the word level, although their performance on tasks of phonetic-phonological processing is not predictive of their comprehension deficit (Blumstein, 1998; Blumstein, Cooper et al., 1977). Additionally, activation is seen in these areas for tasks which involve lexical processing (see Binder & Price, 2001 for review; e.g. Rissman, Eliassen, & Blumstein, 2003).

Scott and Johnsrude argue for an anterior stream for contact between speech sounds and meaning, with basic speech processing occurring in the left anterior temporal lobe, and contact to meaning occurring in inferior frontal areas. This proposal is partially based on a finding that only anterior portions of the left STG are responsive to intelligible speech (Scott, Blank, Rosen, & Wise, 2000). Early neuroimaging results comparing speech stimuli to non-speech control stimuli also implicated areas of the STG and STS both lateral and anterior to Heschl's gyrus in processing speech sounds (Binder et al., 2000; Binder et al., 1997; Demonet et al., 1992; Price, 1998). This view is perhaps less supported by the lesion literature, in which posterior rather than anterior temporal lesions in particular are typically associated with receptive speech deficits.

Binder & Price (2001) propose a one-stream model of speech processing, which has a medial to lateral flow of information, with areas lateral and anterior to Heschl's gyri specifically responsible for sound-to-meaning mappings. They propose a three-stage system of word recognition which proceeds dorsally to ventrally within the STG.

Primary auditory areas are engaged in early acoustic processing of “relatively simple aspects of auditory signals” (204), or pre-phonetic processing. Cortical areas in the superior temporal sulcus and on the lateral surface of the STG, ventral to primary auditory areas, appear to specialize in complex auditory stimuli of all kinds, including the spectral shifts characteristic of speech sounds (Belin et al., 2002; Belin et al., 1999; Belin et al., 2000; Zatorre & Belin, 2001). Even further ventrally, extending beyond the STG into the anterior MTG and inferior temporal lobe are areas implicated in lexical access and semantic processing (Demonet et al., 1992; Price, 1998).

In sum, there are competing views with respect to the role of the temporal lobes in speech processing. Hickock & Poeppel argue that the posterior STG bilaterally is the primary substrate for speech processing, with leftwards lateralization occurring in the context of tasks which require explicit attention to sub-lexical units of speech. Similarly, Scott & Johnsrude argue that this same left posterior STG area should be responsive in explicit phonetic-phonological tasks, but instead argue that the anterior STG should be recruited in receptive language tasks which involve contact with semantic meaning. In the Binder & Price view, the medio-lateral STG should be active in any task which involves processing speech stimuli, irrespective of the task, presumably with increasing leftward lateralization when the task involves contact with a meaningful linguistic representation.   The phonetic categorization task in the following studies presumably involves explicit contact with some sort of phonetic representation, and as such, all models would predict leftward lateralization of activation within the superior temporal gyri. However, the Binder & Price model predicts that such activation should emerge on the medio-lateral surface of the STG, whereas the other two models predict that activation for this task should be in the left posterior STG.

1.3.2. The role of the left inferior frontal gyrus in speech processing

While activation in STG is reported for almost every study involving auditory speech stimuli, activation in frontal areas has been less-consistently reported (Poeppel, 1996a, 1996b). When it is reported, frontal activation is almost universally left-lateralized (but see Paulesu, Frith, & Frackowiak, 1993; Poldrack et al., 1999), often to the exclusion of activation in the right hemisphere. Activation in the L IFG is typically reported when subjects are required either to hold stimuli in memory for a period of time, or required to inspect the stimuli for a particular segment (Demonet 1992, 1994). This activation tends to be adjacent to, and occasionally extending into premotor areas and primary motor cortex. It has been proposed that the left inferior frontal gyrus may be crucial for doing fine-grained acoustic analysis (Zatorre et al., 1992; Zatorre, Meyer, Gjedde, & Evans, 1996), particularly when the analysis involves segmentation of a phoneme from the surrounding phonetic context (Burton, Small, & Blumstein, 2000), and an explicit judgment of phonetic status. As mentioned previously, it has been proposed that the role for this area is in mediating contact between sound-based and articulatory-based phonetic representations (Hickok & Poeppel, 2000; Scott & Johnsrude, 2003).

Patients with Broca's aphasia typically have lesions which include the left inferior frontal gyrus, although the lesions may extend into premotor and motor regions and deep to these regions (H. Damasio, 1998). Both Broca’s and Wernicke’s aphasics may show some impairment on phoneme identification and discrimination tasks, tend to be more impaired in phonetic identification than discrimination tasks, and typically are more impaired in the perception of consonants than vowels. These findings have led some researchers to propose that the basis of aphasic phonological processing disorders stems from a disorder in processing rapid acoustic changes (e.g. Tallal & Piercy, 1975). While the ability of these aphasic patients to process these segments is not normal, the basic structure of the sound category seems to be intact: subjects show no boundary shifts between place of articulation continua for voiced stops (Blumstein, Tartter, Nigro, & Statlender, 1984). Moreover, the asymmetry in performance on phoneme discrimination and identification tasks is suggestive of a disorder in mapping segments onto an abstract linguistic representation, rather than a gross phonological disorder.

Whether the left inferior frontal gyrus is involved in analysis of the fine detail of phonetic structure, segmentation of a phoneme from the surrounding context, or contact with an articulatory representation, all views predict left frontal activation in a phonetic categorization task. What is less clear under each of these views is whether there will be modulation of activation as a function of where the token is along the phonetic continuum is expected.

A more general executive role has also been proposed for the inferior frontal gyri. Thompson-Schill and colleagues (Thompson-Schill, D'Esposito, Aguirre, & Farah, 1997) argue that inferior frontal areas are recruited during tasks which require selection between competing alternatives in language tasks. This hypothesis was formulated with respect to semantic competition, but some evidence exists that this area might also be involved in resolving competition between lexical items with similar phonological structure (Prabhakaran, Blumstein, Myers, & Hutchison, submitted). As such, activation in the left inferior frontal gyrus in the context of a phonetic identification task might reflect competition between similar phonetic categories. An important caveat with respect to the role of this area in language processing is that activation in this area appears under a wide variety of task demands, including those intended to tap syntactic, semantic, and phonological tasks. Therefore, it is as yet unclear whether this area has a specific role for speech perception, or whether it has a general executive function in several aspects of receptive speech and language processing.

1.3.3. Inferior parietal and posterior temporal structures

While the superior temporal gyri and the left inferior frontal gyrus are the areas most commonly implicated in speech processing, other neural areas certainly have a role to play. In particular, inferior parietal areas such as the supramarginal gyrus and angular gyrus, as well as other temporal structures such as the middle temporal gyrus have also been implicated in the mapping of sound to meaning. The supramarginal gyrus, for instance, has been shown to be responsive to competition among words which differ in phonological similarity (Prabhakaran et al., submitted), and patients with lesions in this area have phonological processing deficits (Caplan, Gow, & Makris, 1995). This area has also been implicated in phonological working memory (Paulesu et al., 1993), and in that context is proposed contribute to the neural circuit underlying the phonological loop (Baddeley, 1992). The angular gyrus and middle temporal gyrus have both been implicated in accessing lexical entries (Rissman et al., 2003). Lesion studies also suggest that this posterior temporal/parietal area may be important for naming (Geschwind, 1965; Goodglass, 1993). Grossly speaking, this region seems to be responsive to tasks which tap phonological-lexical processing. It is important to note that activation in these areas is rarely seen for studies which are presumed to tap lower levels of speech and auditory processing, such as phoneme discrimination or passive listening to speech and non-speech stimuli.

1.4. Methodological issues pertaining to fMRI

1.4.1. Lesion studies vs. Neuroimaging

The aphasia literature has been of great importance in guiding the study of the neural basis of sound structure processing. However, some limitations apply to aphasia studies. Because lesions are accidental events, their site and extent are necessarily uncontrollable and variable from patient to patient. Individual variation in level of education, handedness, age, and motivation may affect the recovery of language function after injury, as well as individual differences in cortical organization make it difficult to predict an exact correspondence between structure and function (Varney, Damasio, & Adler, 1989). Moreover, lesions in residual aphasics tend to be large, making it difficult to pinpoint the functional role of particular neural structures.

Neuroimaging techniques complement data from lesion studies, offering researchers an opportunity to study brain activation for phonological processing in normal, healthy populations. While lesion studies in general can tell us a great deal about which brain structures are necessary for a particular cognitive function, neuroimaging data can elucidate which structures are sufficient for performance of that function (Rorden & Karnath, 2004).

1.4.2. Functional Magnetic Resonance Imaging

Although neurophysiological measures such as PET, MEG and ERP have all contributed to our understanding of the neural systems underlying language processing, the studies presented here utilized functional magnetic resonance imaging, or fMRI. As such, issues relating the use of fMRI as a method will be discussed here.

fMRI allows for measuring local changes in blood oxygenation level. The key assumption of fMRI is that regions of increased neural activity bring an excess amount of oxygenated blood to that area (Buckner & Logan, 2001). This change in oxygenated blood is detectable by the MRI, and is known as the Blood Oxygenation Level Dependent (BOLD) signal. MRI has the advantage that, unlike PET, it does not require a radioactive tracer, and is totally non-invasive. While this technique is quite sensitive, the percent signal change is often very small, (2% to 5% for a 1.5 Tesla magnet) so multiple trials are necessary. Traditional fMRI designs require stimulus blocks, but recently techniques have been developed for averaging across a number of single trials (event-related fMRI, or ER-fMRI). This allows experimenters to track changes in hemodynamic response in reaction to a single event, averaged across other events of that type (e.g. Belin et al., 2000). Event-related designs allow for the presentation of multiple stimulus types in a block, which discourages the use of specific strategies, attenuates the effects of expectation, and permits the use of traditional tasks used in cognitive psychology and language processing like lexical decision and phoneme categorization.

The spatial resolution of fMRI is on the order of millimeters, which is superior to other methods like PET and MEG/ERP, but this technique has its limitations as well. Possibly the largest disadvantage associated with using fMRI to study auditory perception in general, and phonological processing specifically is interference due to noise from the scanner. The echo-planar imaging (EPI) sequence used to collect activation data is quite loud (110-130 dB), and can contaminate primary and association auditory cortex data in two ways (Belin et al., 1999). First, the high-amplitude, wide-band frequency scanner noise may induce psychoacoustical masking effects. Second, EPI noise produces significant auditory cortex activity, and can lead to saturation or habituation effects, particularly in primary auditory cortex (A1). The middle two-thirds of Heschl's gyrus is the area most sensitive to scanner noise, with responses as high as a 0.95% increase in blood flow due to scanner excitation (Talavage, Edmister, Ledden, & Weisskoff, 1999). While surrounding association auditory cortex has not been shown to be significantly affected by scanner noise, the increase in blood flow to A1 due to scanner noise could interact with responses in association cortex due to stimulus presentation.

Belin and colleagues (1999; c.f. alsoHall et al., 1999; Talavage et al., 1999) developed a technique called "sparse-sampling" or "stroboscopic presentation" which allows for stimulus presentation during silent periods between scans. By imaging just a few slices, they were able to reduce their volume acquisition time to one second, which was repeated every ten seconds, with a nine second silent period between scans. Stimuli were presented at variable delays after each scan. In addition to allowing for stimulus presentation in relative silence, this technique permits sampling of the time-course of the hemodynamic response after stimulus presentation. Some limitations of this design are that it is suitable only for fairly short stimuli (2-3 seconds), and that experimenters are generally required to specify a narrower imaging window than usual, so as to decrease the acquisition time. Be that as it may, this method is well suited for studying the finer details of acoustic-phonetic processing that are planned in the current studies.

In addition to noise artifacts, fMRI is also susceptible to other artifacts of scanning. fMRI is more sensitive to movement artifacts than PET, which can result in spurious (often extra-cranial) activation (Buckner & Logan, 2001). If movement is by chance correlated with the condition or task, spurious activation will be falsely attributed to task demands. Motion correction algorithms can eliminate the effects of movement artifacts to a large degree. Of particular importance to auditory processing studies is the fact that signal dropout can occur at sites near the sinuses, due to the water/air interface in these regions. Unfortunately, the anterior temporal lobes are very near the sinuses, and therefore especially prone to this kind of artifact. Because the signal-to-noise ratio in this region is so low, null results are not necessarily indicative of a lack of change in brain activity (Buckner & Logan, 2001).

1.5. Specific goals of three neuroimaging experiments

Three experiments were conducted in order to investigate the neural structures underlying phonetic categorization. In all studies, subjects performed a phonetic categorization task on auditory stimuli which ranged along a VOT continuum while activation levels were measured using fMRI. All studies used a sparse sampling design (Belin et al., 1999; Hall et al., 1999) in which stimuli were presented during the pauses between functional scans. This allowed for the presentation of speech stimuli in relative silence, and as such allowed us to examine neural responsiveness to fine differences in acoustic structure for stimuli along the VOT continuum. In all experiments, the statistical comparisons of interest were between speech stimuli along the continuum, rather than between these stimuli and control stimuli (e.g. tones). This allowed for a relative degree of consistency in the nature and difficulty of the task, and maximized the probability that differences observed in activation patterns to stimuli along the continuum were due to factors associated with the phonetic processing of the stimuli, rather than artifacts of the task itself.

Experiment 1 was designed to elucidate the role of both posterior and anterior peri-sylvian areas in the phonetic categorization process. The specific question addressed was whether graded patterns of neural activation would emerge to stimuli which varied in their position on the VOT continuum at various sites along the neural processing stream. Participants made a phonetic categorization decision (/da/ or /ta/) on synthetic speech stimuli which ranged along a VOT continuum from 0 msec VOT (/da/) to 40 msec VOT (/ta/) in 10 msec steps. Activation levels were measured for endpoint stimuli which were good exemplars of their phonetic categories, to exemplars which were closer to the phonetic boundary, and to stimuli which fell on the phonetic category boundary itself. It was hypothesized that graded activation would be observed in both frontal and posterior areas, potentially with greater activation for stimuli near the phonetic boundary, as those stimuli should require more processing to resolve their phonetic category status.

Experiment 2 was designed to separate the potential effects of acoustic-phonetic competition and token ‘goodness of fit’ on activation in the superior temporal gyri and the left inferior frontal gyrus. To this end, participants made a phonetic categorization decision (/da/ or /ta/) on synthetic speech stimuli which were either poorer exemplars of the phonetic category and close to the phonetic category boundary, good exemplars from the center of the phonetic category, or poorer exemplars of the phonetic category and far in acoustic space from the boundary. Competition between phonetic categories is expected to increase as stimuli approach a contrasting phonetic category, while goodness of fit decreases as stimuli become further from the phonetic category center. Given the postulated role of inferior frontal areas in mediating competition (Thompson-Schill et al., 1997), it was hypothesized that inferior frontal areas should be maximally responsive to the effects of competition between phonetic categories. In contrast, the superior temporal gyri have been shown to be sensitive to phonetic category structure (Guenther, Nieto-Castanon, Ghosh, & Tourville, 2004). As such, it was hypothesized that the STG bilaterally would show sensitivity to the goodness of fit of tokens to their phonetic category.

Experiment 3 was designed to explore the effects of top-down processing on phonetic categorization, and in particular the effects lexical status on the phonetic categorization process. Participants performed a phonetic categorization task on stimuli which ranged along a VOT continuum (/g/ to /k/). Two continua were presented in which the stimuli had a real word at one end of the continuum (i.e. ‘gift’, ‘kiss’) and a non-word at the other end of the continuum (i.e. ‘kift’ , ‘giss’). As has been shown in previous behavioral experiments (Burton & Blumstein, 1995; Ganong, 1980; Miller & Dexter, 1988; Pitt, 1995) the presence of a real word on one end of a phonetic continuum is expected to shift the phonetic category boundary in a phonetic categorization task such that there are more responses (in the case of the gift—kift continuum, /g/ responses) which result in a real word (in this case, ‘gift’). The activation patterns observed for a given VOT value near the phonetic category boundary are expected to vary as a function of where the phonetic boundary actually falls in the continuum of interest. The emergence of differences in activation in either sensory (STG) or executive (IFG, anterior cingulate) areas will bring evidence to bear on the nature of lexical effects as either sensory-level or decision-related effects. Specifically, the emergence of differences in activation for the same VOT value in the STG is suggestive that lexical effects operate at the perceptual level. In contrast, modulation of activation in the IFG or anterior cingulate (in the absence of STG modulation) may be taken as evidence that the lexical effect operates at decision stages of processing.

Taken together, these experiments are designed to shed light on both the neural bases of phonetic categorization processes as well as the structure of the cognitive architecture which mediates the phonetic categorization process.

2. CHAPTER 2: EXPERIMENT 1, CATEGORIZATION OF VOICE ONSET TIME[1]

2.1. Introduction

Most models of speech perception assume multiple levels of processing between the auditory input and the mapping of sound structure to meaning. In this view, the auditory input from the peripheral auditory system is transformed from simple acoustic features of frequency and amplitude to more generalized auditory patterns or properties. These spectral-temporal patterns are in turn converted to a more abstract category representation corresponding to the phonetic categories of speech. It is generally assumed that lexical entries are represented in terms of these categories.

The speech perception process is complicated by variability in naturally produced speech which has large consequences on the acoustic output. A number of factors influence the production of a given phoneme, including speech rate, phonetic context, dialect, and even properties of the lexicon. Some variability is inherent in the production system simply because speakers do not have perfect articulatory control. For instance, the acoustic parameter known as voice-onset time (VOT) is one cue which distinguishes stop consonants which contrast in the phonetic feature of voicing (Lisker & Abramson, 1964). Within a given phonetic category, for instance, /t/, a speaker may produce VOTs ranging from 40 msec VOT to 170 msec VOT. Therefore, listeners need some mechanism by which to attend to the relevant acoustic-phonetic features of the input in order to ultimately match the range of VOT values to a common phonetic category.

Listeners show both all-or-none (categorical) as well as graded responses to

---

[1] This chapter is adapted from Blumstein, Myers & Rissman (2005) in press, Journal of Cognitive Neuroscience

speech stimuli which vary along a voicing continuum. In particular, listeners tend to be at chance in discriminating between tokens which fall within the same phonetic category, yet relatively good at discriminating tokens which fall in two distinct phonetic categories (Liberman et al., 1961). The source of poor within-category discrimination has been extensively debated in the literature. Kuhl (1991) claims that within-category insensitivity arises as a function of a learned warping of the acoustic space near the phonetic category prototype. Likewise, Guenther and colleagues (Guenther & Gjaja, 1996; Guenther, Husain, Cohen, & Shinn-Cunningham, 1999) have proposed that this learned warping is a natural consequence of phonetic categorization, and have shown that categorization training decreases the discriminability of within-category contrasts, whereas discrimination training increases sensitivity to such contrasts. Others (Steinschneider, Schroeder, Arezzo, & Vaughan, 1994, 1995) have instead proposed that within-category insensitivity for voicing contrasts specifically is biologically determined, and arises as a function of the response characteristic of a set of neurons found in primate auditory cortex.

While subjects are less sensitive to within-category than between-category distinctions, clear evidence exists for sensitivity to within-category distinctions. Within a phonetic category, listeners show graded sensitivity to the ‘goodness of fit’ of an exemplar to its phonetic category. For instance, although listeners judge both good exemplar voiceless stop consonants and voiceless stop consonants with reduced VOTs as belonging to the same phonetic category, they show longer reaction times in discrimination tasks to such stimulus pairs than to acoustically identical pairs (Pisoni & Tash, 1974). They also show longer response latencies for stimuli with reduced VOTs in a phonetic categorization task (Andruski et al., 1994), and they rate these stimuli as poorer exemplars of the voiceless phonetic category (Miller & Volaitis, 1989). Recent evidence from eye-tracking indicates that subjects show graded effects of within-category variation on eye-movements to targets like ‘bear’ and ‘pear’ (McMurray et al., 2002; McMurray et al., 2003). This suggests that listeners do have access to information about within-category variability of the signal. Such information may be shown to be preserved at higher levels of the speech processing system. For instance, Andruski et al (Andruski et al., 1994) showed that ‘cat’ primes ‘dog’ to a lesser extent when the initial phoneme of that word is modified to make it a poorer exemplar of a /k/ sound. This indicates that within-category phonetic variation is not only perceptible to the listener, but may have down-stream consequences at higher levels of language processing. Given that within-category variation may be informative, providing information about rate, phonetic context, or speaker-specific characteristics, sensitivity to within-category variability may be advantageous to the listener at early stages of processing.

In this study, the neural bases of phonetic categorization were examined as subjects performed a phonetic categorization task on stimuli which ranged along a VOT continuum in the fMRI scanner. This paradigm offers insight into the brain basis of phonetic categorization systems generally, and of the perception of the phonetic feature of voicing specifically. Finally, the brain mechanisms responsible for graded and categorical responses to speech stimuli can be investigated.

Evidence from both the aphasia and neuroimaging literatures supports the notion that the speech perception system consists of a network of primarily left-lateralized perisylvian areas including the left inferior frontal gyrus, the superior temporal lobes bilaterally, and left inferior parietal areas such as the angular and supramarginal gyri. Both Broca's aphasics, who typically have damage to left frontal structures, and Wernicke's aphasics, who typically have damage to left temporal and/or parietal structures, have deficits in discriminating phonetic contrasts such as *pill* vs. *bill* (Blumstein, Baker et al., 1977), and also perceiving acoustic-phonetic properties distinguishing phonetic categories such as voice-onset time or formant transitions (Blumstein, Cooper et al., 1977; Blumstein et al., 1984). The distributed nature of the speech perception system is also reflected in the neuroimaging literature by widespread activation of left hemisphere perisylvian areas in a variety of speech perception tasks.

The bilateral temporal lobes are widely believed to subserve most basic acoustic analysis of speech. Given that the primary auditory area, Heschl's gyrus, surfaces in the temporal lobe, it has traditionally been assumed that speech perception recruits posterior and, in particular, temporal lobe structures. Early evidence from the aphasias supports this view. Wernicke's aphasics with damage in left temporal structures displayed severe auditory comprehension deficits (Geschwind, 1965). However, pure word deaf patients are the only aphasic group for which impairments appear to be limited to speech perception (Auerbach et al., 1982; Engelien & Huber, 2001; Poeppel, 2001). These patients typically have either bilateral temporal lobe lesions or lesions that involve the left auditory cortex and extend deep, cutting off right hemisphere auditory input. Neuroimaging studies support these findings from the aphasias. The superior temporal gyri (STG) and superior temporal sulci (STS) show activation for speech as well as for complex auditory non-speech stimuli (Binder et al., 2000; Binder et al., 1994). The STG also show preferential activation to human vocal sounds in general, and the acoustic phonetic patterns which distinguish speech specifically (Belin et al., 2002; Belin et al., 2000; Scott et al., 2000). In sum, the primary auditory cortex and auditory association areas (STG and STS) bilaterally have been implicated in early stages of acoustic-phonetic processing (Belin et al., 1999; Binder et al., 2000; Price et al., 1992; Scott & Wise, 2004; Zatorre et al., 1996).

Superior temporal areas have also been implicated in both graded and categorical responses to variation in phonetic category. Guenther et al. (2004) showed decreased activation in several superior temporal areas when subjects listened to a good exemplar of the phonetic category /i/, compared to when they listened to an /i/ token that was closer to the phonetic category boundary with /I/. This sensitivity to within-category differences suggests that the temporal lobes may be sensitive to the “goodness of fit” of a token to its phonetic category.

Research using ERP and MEG has also shown that electrophysiological signatures which are believed to arise from the left temporal lobe are sensitive to both graded and categorical aspects of phonetic category structure (see Phillips, 2001 for review). Specifically, Sharma and colleagues (Sharma & Dorman, 1999; Sharma & Dorman, 2000; Sharma, Marsh, & Dorman, 2000) showed that the auditory mismatch negativity, or MMN, was much larger when subjects heard a native-language phonetic category contrast, than for native language within-category contrasts or non-native contrasts. Similarly, Phillips and colleagues (Phillips et al., 2000) demonstrated the presence of a MMF (the MEG analogue of the MMN) for oddball stimuli which crossed phonological boundaries, but not those which did not. The MMF signature in this study is believed to arise from the left superior temporal gyrus. The authors conclude that this signature is related to a phonetic-phonological level of processing at which much of the acoustic information initially available in the signal has been discarded, and at which information about native-language contrasts is accessed.

In contrast, a second auditory signature, the N1 (also known as the N100) has shown sensitivity to within-category changes in VOT. The N100 has been linked to any detection of an acoustic event with a distinct onset (Naatanen & Picton, 1987). When subjects were presented with stimuli with VOTs of 30 msec or less, which typically corresponds to the voiced end of the continuum, one N100 peak was observed, which had a latency which correlated with the VOT of the stimulus. In contrast, when subjects heard a stimulus with a VOT greater than 40 msec (usually voiceless), two N100 peaks emerged, the first, which was timelocked to the onset of the burst, and a second which correlated with the onset of voicing (Sharma & Dorman, 1999). This response did not change as a function of where the voicing boundary actually fell (Sharma et al., 2000) or as a function of individual language experience (Sharma & Dorman, 2000). N100 signatures were equivalent for English and Hindi speakers for an acoustic phonetic continuum which contained two Hindi phonetic categories but only one English category (Sharma & Dorman, 2000). The N100 is therefore thought to reflect the basic acoustic analysis of speech, without regard to the specifics of the native language phonology.

In sum, there is reason to suspect that the superior temporal lobes have the capacity to detect within-category variation in voice-onset time. As such, graded activation in the superior temporal lobes might be expected as stimuli vary in their ‘goodness of fit’ to a phonetic category. However, given the existence of categorical MMN responses which are also believed to arise from the superior temporal lobes, it is an open question as to whether such activation will emerge.

While the superior temporal lobes may be the primary substrate for the processing of speech information, neuroimaging studies have shown activation of anterior as well as posterior structures of the left hemisphere (Burton, 2001; Zatorre, Belin, & Penhune, 2002; Zatorre et al., 1996). The inferior frontal gyri seem to be particularly responsive to task demands and presumably the different cognitive mechanisms and levels of speech processing required by different speech tasks (Poeppel, 1996a). Anterior structures such as the left inferior frontal gyrus, and to a lesser extent the middle frontal gyrus have been implicated in tasks that involve overt segmentation of the speech signal (Burton et al., 2000; Zatorre et al., 1996). Portions of the left inferior frontal gyrus have also been implicated in phonological processing generally (Poldrack et al., 2001). A series of studies by Thompson-Schill has also implicated the left IFG in selection between competing stimuli (Thompson-Schill et al., 1997). In a phonetic categorization task, stimuli which are closer to the phonetic boundary may be subject to a greater degree of competition from the contrasting phonetic category. If this is the case, modulation of activation may emerge in left inferior frontal areas as a function of the proximity of a stimulus to the phonetic category boundary.

Posterior temporo-parietal areas such as the angular gyrus and supramarginal gyrus have also been implicated in the processing, storage, and retrieval of phonological information. These areas have been shown to be involved as subjects access language-specific phonological information (Jacquemot, Pallier, LeBihan, Dehaene, & Dupoux, 2003), in the retrieval and manipulation of information in phonological working memory (Awh et al., 1996; Jonides et al., 1998; Paulesu et al., 1993), and in resolving competition between lexical items with similar phonological structure (Prabhakaran et al., submitted) Evidence seems to suggest a role for temporo-parietal areas in the processing of higher-level phonological information rather than the specific acoustic phonetic details of speech. As such, this area may not show sensitivity to within-category variation among stimuli, and may instead show ‘categorical’ responses to speech stimuli.

It is the goal of the current study to examine the neural systems underlying the perception of the phonetic categories of speech. To this end, we will investigate the perception of voice-onset time, a temporal cue that distinguishes initial voiced, e.g. [d], from voiceless consonants, e.g. [t], in English using a phonetic categorization task. A synthetic voice-onset time continuum will be used ranging from 0 to 40 ms VOT in 10 ms steps, and subjects will be asked to categorize the stimuli as either [d] or [t]. We will examine potential differences in neural activation patterns in the perception of a speech continuum containing stimuli which are good exemplars of a phonetic category (endpoint stimuli, VOT 0 and 40), stimuli that are perceived as members of the phonetic category but are near to the acoustic-phonetic boundary (within category stimuli, VOT 10 and 30), and a stimulus that is inconsistently perceived as a member of either category (boundary value stimulus, VOT 20). In addition, there will be a tone control task in which subjects will be required to categorize the stimuli as either high tone or low tone.

It is hypothesized that the superior temporal gyrus should show extensive bilateral activation for the processing of a VOT continuum. What is less clear is whether this region will be responsive to differences among the different acoustic exemplars of the continuum. It is possible that the role of the superior temporal gyrus is to do an early acoustic analysis of the sensory information inherent in the stimuli, and to pass this information to other brain regions for higher levels of processing (Binder et al., 2000; Scott & Wise, 2004). In such a case, although there would be extensive activation of the STG bilaterally, there would not be modulation of activation as a function of the stimulus type. As stated previously, evidence using the MMN paradigm indicates that the temporal lobes are sensitive to phonetic category status. Therefore, it is possible that the STG will show sensitivity to the acoustic-phonetic differences among the stimuli and hence will show modulation of activation. In particular, graded patterns of activation may emerge in this area as a function of the goodness of fit of a stimulus to its phonetic category.

Because the phonetic categorization task is a linguistic one, it is expected that the activation patterns of areas other than the superior temporal gyrus will be left-lateralized. In particular, there should be activation of both left frontal and temporo-parietal areas. Nonetheless, it is hypothesized that the patterns of activation in these areas will reflect different aspects of the phonetic categorization process. In particular, because frontal areas appear to be involved in decision processes and are sensitive to the difficulty of the decision, it is hypothesized that areas such as the left inferior frontal gyrus and cingulate should show modulation of activation across the VOT continuum as a function of the goodness of fit of the stimulus to its phonetic category. Thus, increasingly greater activation should occur for the within category stimuli and the boundary value stimulus compared to the endpoint stimulus, reflecting the increasing resources needed to make a phonetic category decision for these tokens.

It is also the case that the within category stimuli are nearer to the phonetic boundary and closer in acoustic space to the contrasting phonetic category. As such, these stimuli are not only more difficult to map on to a phonetic category, but there is also greater competition between the voiced and voiceless phonetic categories. Such increased competition may also potentially recruit frontal areas such as the inferior frontal gyrus (Thompson-Schill et al., 1997).

As described earlier, the phonetic categorization process requires that listeners categorize different exemplars as members of the same phonetic category. There are several processes involved in determining such phonetic category membership. These include mapping of the sound structure to its phonetic representation, and classifying the representation as a member of a particular phonological category. Given the role of inferior parietal areas such as the angular gyrus and supramarginal gyrus in phonological processing, these areas might demonstrate activation which reflects the categorical nature of phonological representations. In particular, activation differences might be seen between boundary stimuli and other stimulus types, but no differences between within-category and endpoint stimuli which fall within the same phonetic category.

In addition, subjects performing this task are also likely accessing the name of the stimulus as it is categorized. Naming and lexical processing have been linked to temporo-parietal areas such as the left middle temporal gyrus (Hickok & Poeppel, 2004; Scott & Wise, 2004) and the angular gyrus (Geschwind, 1965). These areas in particular may instead show the greatest activation to those tokens that correspond to the ‘best fit’ to that phonetic category.

2.2. Methods & Materials

2.2.1. Participants

Thirteen adult native speakers of English (11 females) between the ages of 20 and

59 (mean age=25.5 yrs ± 10.5 yrs) participated in the study each screened for a history of neurological disease or injury. All reported normal hearing and were right-handed as determined by Oldfield's handedness inventory (Oldfield, 1971). Participants gave written informed consent according to guidelines established and approved by the Human Subjects Committees of Brown University and Memorial Hospital of Rhode Island, and were screened for MR safety before entering the scanner. Participants received modest monetary compensation for their time. One subject was excluded from both behavioral and functional analyses due to improper image acquisition parameters.

2.2.2. Materials

Stimuli consisted of five synthetic speech stimuli taken from a larger continuum ranging from [da] to [ta] synthesized at Haskins Laboratory using a parallel synthesizer. The five stimuli ranged in voice-onset time (VOT) from 0 msec ([da]) to 40 msec ([ta]) in 10 msec steps. (Figure 1). VOT was manipulated by replacing the periodic source with an aperiodic source in ten msec increments starting from the stimulus onset. Two sine-wave tones with frequencies of 910 Hz ('Low Tone') and 1320 Hz ('High Tone') were also generated for use in the tone task. All stimuli were 230 msec in length.

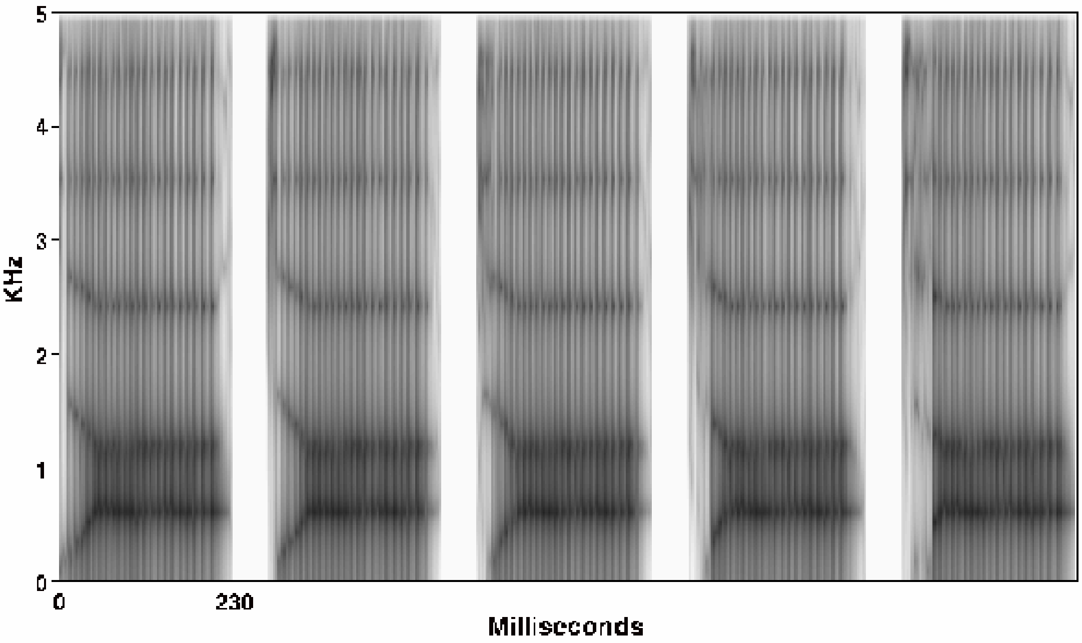

Figure 1. Wide-band spectrograms of speech stimuli used in Experiment 1, ranging in VOT from 0 msec to 40 msec VOT (left to right). Each stimulus consisted of a five-formant pattern. The onset frequencies were 200 Hz (F1), 1350 Hz (F2), and 3100 Hz (F3). Formant transitions into the vowel were 40 ms to a steady-state vowel with formant frequencies of 720, 1250, and 2500 Hz, respectively. F4 and F5 remained steady throughout at 3600 and 4500 Hz.

2.2.3. Behavioral Procedure

The experiment consisted of two runs in each of two tasks, a phonetic categorization task (PC), and a tone categorization task (TC). Participants received the four runs in a fixed order (PC, TC, PC, TC). Each PC run consisted of 20 repetitions of each of the five synthesized syllables, and each TC run consisted of 20 repetitions of each of the two tone stimuli (see *MR Imaging* for details of stimulus timing). Stimuli within each run were presented in a fixed, pseudo-randomized order. During scanning, participants listened to stimuli through MR compatible headphones (Resonance

Technology™, Northridge,California) set at the maximum comfortable volume, and indicated responses via two buttons of a MR-compatible button box placed by the right hand (Resonance Technology™). Stimuli were presented using an IBM ThinkPad running the AVRunner program, designed for accurate timing of auditory stimuli presentation and collection of reaction time (RT) data (Mertus, 1989 http://www.cog.brown.edu/localSites/mertus/BlissHome.htm).

For the PC task, participants were instructed to listen to each syllable and to decide whether the syllable was [da] or [ta], by pressing a corresponding button as quickly and accurately as possible. For the TC task, subjects were told to listen to each tone stimulus, and to decide whether it was the ‘high’ tone (1320 Hz) or the ‘low’ tone (910 Hz), by pressing the corresponding button. The mapping of the button (left or right) to the response type (‘ta’ or ‘da’; ‘high’ or ‘low’) was counterbalanced across subjects. Participants were given five trials of each task for practice during acquisition of the anatomical dataset. RT and categorization data were collected for all subjects; RTs were measured from the onset of the stimulus.

2.2.4. MR Imaging

Whole brain MRI was performed with a 1.5 T Symphony Magnetom MR system (Siemens Medical Systems, Erlangen, Germany) equipped with echoplanar imaging capabilities. Each participant’s head was aligned to the magnetic field center. Participants were instructed to refrain from moving the head during MR imaging, and were reminded to keep their eyes closed.

For anatomical co-registration, we acquired a 3D T1-weighted magnetization prepared rapid acquisition gradient echo (MPRAGE) sequence (TR=1900 msec, TE=4.15

msec, TI=1100 msec, 1 mm^3 isotropic voxel size, 256 X 256 matrix) reconstructed into 160 slices. A multi-slice, ascending, interleaved, echo planar imaging (EPI) sequence with 15 axial slices constituted the functional scans (5 mm thickness, $3mm^2$ axial in-plane resolution, 64 X 64 matrix, 192 mm^2 FOV, FA = 90°, TE=38 msec, TR=2000 msec). Before EPI images were acquired, the center of the imaged slab was aligned to each participant's corpus callosum using a sagittal localizer image, which allowed for the collection of functional data from bilateral peri-sylvian cortex. A functional acquisition sequence was used in which auditory stimuli were presented during silent gaps between volume acquisitions (Belin et al., 1999; Hall et al., 1999; Jancke, Wustenberg, Scheich, & Heinze, 2002). Each EPI volume acquisition was obtained in 1200 msec (80 msec per slice) followed by 800 msec of silence, during which the auditory stimulus was presented (Figure 2), yielding an effective volume repetition time of 2000 msec.

We used an event-related design presenting stimuli at different equally probable trial onset asynchronies (TOA = 2, 4, 6, 8, 10 s). Four volumes were acquired prior to the onset of the first stimulus to avoid contamination of functional data by T1 saturation. These four volumes were eliminated from further analysis. Each of the two PC runs consisted of 304 echo-planar volumes, and each TC run consisted of 124 volumes, for a total of 856 EPI volumes.

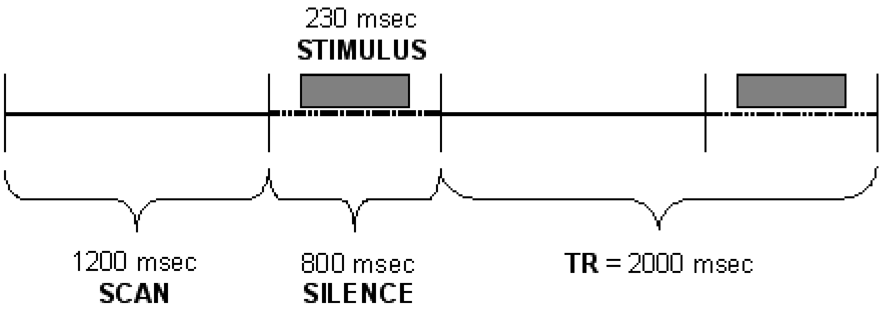

Figure 2. Scanning schematic.

2.2.5. Data Analysis

2.2.5.1. Behavioral Analysis

RT and categorization data were collected for all participants. Reaction time means were computed for each VOT stimulus for each subject after eliminating responses having RT > 2 sec or > 2 SD from the mean for a particular stimulus for that subject. The percentage of ‘da’ responses was also calculated for each subject for each VOT value. Repeated measures ANOVAs were performed separately on the RT means and on the percent ‘da’ values for each subject.

2.2.5.2. MR Analysis

2.2.5.2.1. Image Processing

All analysis of imaging data was carried out using AFNI (Cox & Hyde, 1997). Functional MR images were corrected for head motion by aligning all volumes to the fourth collected volume using a six-parameter rigid body transform (Cox & Jesmanowicz, 1999) and then resampled to 3mm isotropic voxels and warped to Talairach and Tournoux space (Talairach & Tournoux, 1988) using AFNI tools. The images were then spatially smoothed with a 6 mm full width half maximum Gaussian kernel. These pre-processed datasets became the input to a deconvolution analysis.

2.2.5.2.2. Statistical Analysis

Each subject's EPI data were submitted to deconvolution analysis to estimate the individual hemodynamic response during each stimulus condition. A reference time series was created for each of the seven trial types (five VOT conditions, two tone conditions) by convolving the stimulus presentation times with a stereotypic gamma-variate hemodynamic response curve provided by AFNI (Cohen, 1997). The resulting seven reference waveforms were used as covariates in the deconvolution analysis. The six output parameters of the motion correction analysis (x, y, and z translations, roll, pitch and yaw) were also included as covariates in order to remove residual motion artifacts from the EPI time series. These thirteen reference functions were then submitted to AFNI's 3dDeconvolve program, which output raw fit coefficients for each subject for each condition on a voxel-by-voxel basis. The fit coefficients were then converted to percent change by dividing the coefficient from each voxel by the baseline for that voxel, which was calculated by computing an experiment-wise mean for that voxel.

The percent change values for each condition and each participant were entered into a two-way, mixed-factor ANOVA using stimulus condition as the fixed factor and participant as the random factor. A group statistical map was created for each stimulus condition. In order to reduce the number of statistical comparisons, the five VOT conditions were grouped into three stimulus categories: the 0 msec and 40 msec tokens constituted the 'endpoint' category, the 10 msec and 30 msec tokens constituted the 'within category' stimuli, and the 20 msec VOT stimulus was referred to as the 'boundary' stimulus. Four planned comparisons were carried out in the same analysis:

endpoint vs. within category, endpoint vs. boundary, within category vs. boundary, and endpoint vs. tone stimuli. Using standard AFNI methods, we used a voxel-level threshold of $p<0.025$ and a cluster threshold of 81 contiguous $3mm^3$ voxels in Talairach space which, based on Monte Carlo simulations, yielded a corrected threshold of $p<0.01$.

In order to examine the time course of the hemodynamic response within those regions identified in the previous analyses, a second deconvolution analysis was performed in which no a priori assumptions were made about the shape or temporal properties of the hemodynamic function. For this analysis, a binary stimulus function was created for each condition, with ones at time points when stimuli were presented and zeroes elsewhere. The deconvolution was carried out for each subject as described above, with the binary stimulus functions substituting for the convolved stimulus functions. Voxel-wise scaling coefficients were output at two-second intervals ranging from 0-20 seconds post-stimulus, and coefficients were converted to percent signal change values as described in the gamma-based deconvolution analysis. Time series graphs were generated for each subject by taking the means of all activated voxels within a cluster at a range of post-stimulus time lags, and these were then averaged to create a group time series graph for the cluster

Region of interest (ROI) analyses were carried out in Heschl’s gyrus and the superior temporal gyrus bilaterally because of a priori hypotheses predicting activation differences in primary auditory and auditory association areas as a function of stimulus type. These areas were defined based on rough anatomical localization masks provided by AFNI (Lancaster et al., 2000). Additionally, in order to investigate the hypothesis that functional divisions exist along the anterior to posterior extent of the left STG, the left STG mask was divided into approximate thirds, with the anterior third bounded posteriorly by a plane at y = -8, the middle third extending between planes at y = -8 and y = -35, and the posterior third bounded anteriorly by a plane at y= -36. The mean percent signal change value from the gamma-based deconvolution analysis was taken of all voxels for each subject and condition within an area defined by the rough anatomical masks provided by AFNI. Only voxels which were imaged in all subjects and also fell within the anatomical area of interest contributed to the mean. By-condition and by-subject means were entered into separate within-subject repeated-measures ANOVAs for each anatomical area, with stimulus condition (endpoint, within category, or boundary) as the variable of interest.

A regression analysis was performed to investigate the possibility that some portion of the activation differences revealed in the first ANOVA might be attributable to differences in the difficulty of the task across stimuli (as estimated by reaction time) rather than differences in the processing of the stimuli per se. This analysis was designed to identify those areas in which percent signal change correlated significantly with reaction time. Mean by-condition RT for each subject, the squares of these means, and the log of those means served were the regressors of interest, which were respectively meant to identify regions responsive to any linear, quadratic, or logarithmic relationship between activation and reaction time. Additional regressors for each subject and condition were included to factor out variability attributable to those variables. In this analysis, each VOT stimulus type was entered into the regression separately in order to assess the contribution of reaction time to individual variation among the five VOT stimulus types. The employed model tested what portion of the variance could be absorbed by the RT variables which was not already accounted for by the subject and condition variables. The resultant statistical maps were thresholded as above. Tone stimuli were not included in the analysis, as there was no reason to expect any variability in the difficulty of the two tone types, ‘High’ and ‘Low.’ The behavioral data support this conclusion as there was no significant difference in reaction time (one-tailed $t=1.058$, $p<0.158$) or in accuracy (one-tailed $t=0.919$, $p<0.189$) between High and Low tones, as indicated by paired t-tests.

A final analysis was carried out to identify those regions where there was a significant correlation between activation and a “Goodness of Fit” regressor, factoring out RT, subject and condition variables. To test this hypothesis, we used only the VOT stimuli in a linear regression analysis having one ‘goodness of fit’ regressor which assumed a linear increase in activation as stimuli approached the phonetic boundary (endpoint stimuli--0 msec and 40 msec VOT=1, within category stimuli--10 msec and 30 msec VOT=2, boundary stimuli--20 msec VOT=3), and one regressor for each subject. As in the RT regression, mean by-condition RT for each subject, the squares of these means, and the log of those means were included in order to factor out activation which was associated with these variables. The resultant statistical maps were thresholded as indicated above.

2.3. Results

2.3.1. Behavioral Results

The behavioral results for both reaction time and categorization data for VOT stimuli are shown in Figure 3, and mean and standard deviation values for all stimuli are shown in Table 1. Subjects exhibited the classical categorical perception identification function. A one-way repeated-measures ANOVA using the percentage of 'da' categorizations for each subject for each stimulus on the continuum was significant (F(4,44)=133.7; p<0.001). Post hoc Newman-Keuls tests revealed significant differences in percent 'da' responses between the 10 msec and 20 msec VOT stimuli and between the 20 msec and 30 msec VOT categories at p<0.05. There was no significant difference between the 0 and 10 msec stimuli, nor was there a significant difference between the 30 and 40 msec stimuli, indicating that within-category shifts of voice-onset time have little effect on phonetic categorization decisions. In contrast, the reaction time results indicate sensitivity to within-category shifts in VOT, with reaction times increasing as the stimuli approach the phonetic boundary. A repeated-measures ANOVA indicated a main effect of VOT on reaction time, (F(4,44)=6.503; p<0.001). Post-hoc Newman-Keuls tests indicated no significant differences between any individual VOT stimulus and its VOT neighbors. However, when stimuli were grouped into endpoint (0 and 40 msec VOT), within category (10 and 30 msec VOT), and boundary (20 msec VOT) categories as was done in the fMRI analysis, post-hoc tests indicated significant differences at p<0.05 between all stimulus categories.

Table 1. Behavioral data collected from 12 participants in the MRI system. The first column indicates the mean percent 'da' responses for the VOT stimuli, or the mean percent correct for the Tone stimuli. The second column indicates the mean reaction time. Standard error of the mean is indicated in parentheses.

Stimulus	**% 'da' responses** (VOT) or **% correct** (Tones)	**RT in msec**
0 msec VOT	95 % *(2.26 %)*	680 *(34)*
10 msec VOT	87 % *(4.48 %)*	724 *(45)*
20 msec VOT	47 % *(8.59 %)*	782 *(45)*
30 msec VOT	0 % *(0.00 %)*	748 *(47)*
40 msec VOT	0 % *(0.00 %)*	657 *(39)*
High Tone	96 % *(1.20 %)*	646 *(29)*
Low Tone	98 % *(1.07 %)*	633 *(35)*

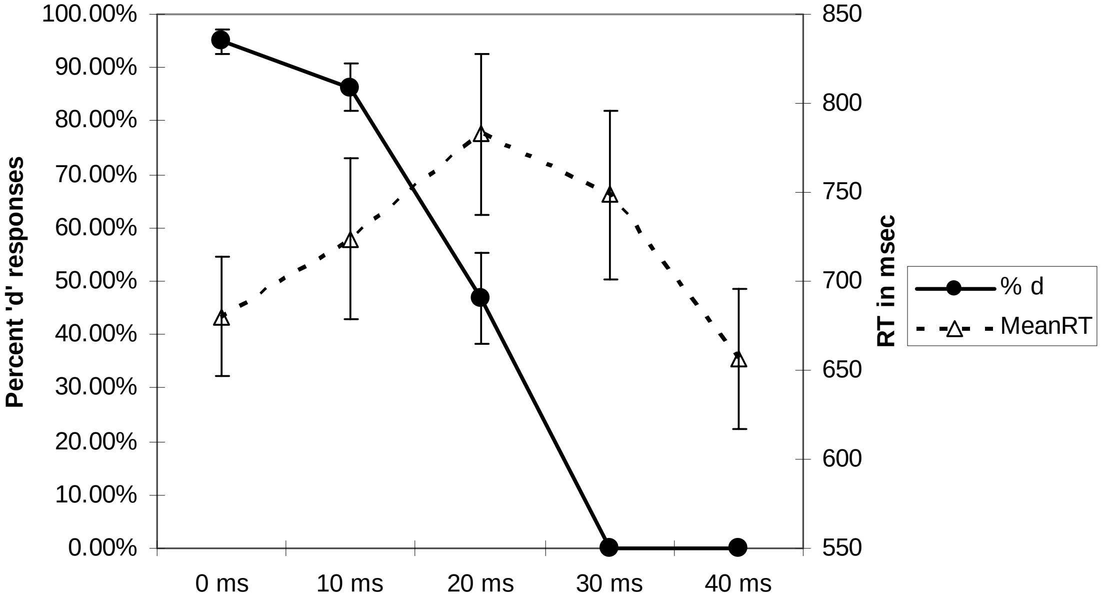

Figure 3. Mean RT in msec (triangles) and mean % ‘da’ responses (circles) for the 12 MR participants. Bars represent standard error of the mean.

2.3.2. Functional results

2.3.2.1. Mean activation analysis

In order to qualitatively compare the extent of activation across conditions, mean activation maps were generated for all stimulus categories. Figure 4 shows a representative stimulus from the endpoint (0 msec), within category (10 msec), and tones (high tone) categories together with the boundary stimulus (20 msec). In general, all syllable conditions yielded extensive bilateral activation centering in peri-sylvian and midline regions. Activation tended to be of greater intensity and extent in left hemisphere regions than in the homologous right hemisphere regions. In addition, a greater amount of activation was observed for the boundary value stimuli than the within category stimuli, and for the within category stimuli than for the endpoint stimuli. The tone stimuli activated regions similar to those activated in the syllable condition, but both activation intensity and extent of activation was less lateralized.

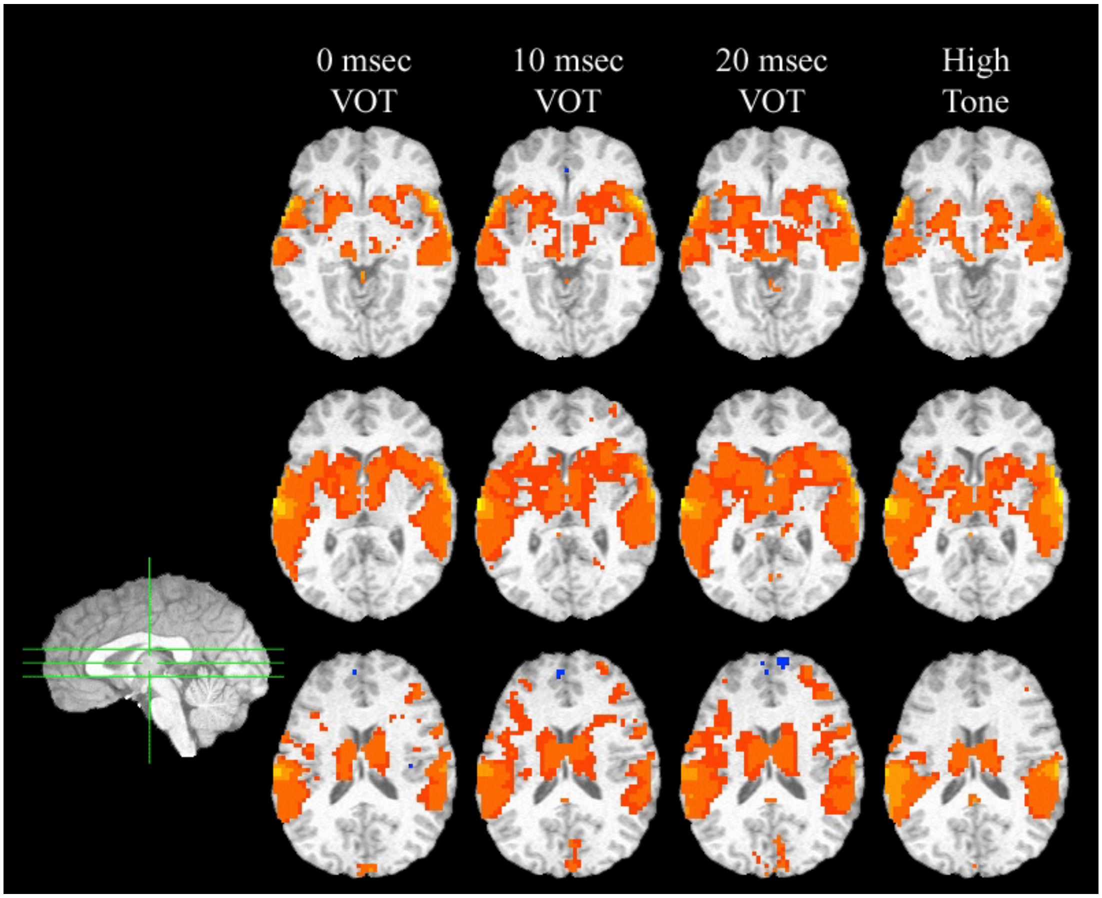

Figure 4. Group activation maps, thresholded a voxel-wise $p<4.0x10^{-5}$, uncorrected. Axial slices at z=-2, z=8, and z=18, bottom to top.

2.3.2.2. Cluster analysis

A summary of all clusters activated in the planned comparisons significant at a corrected p<0.01 is shown in Table 2. The following discussion will focus on the planned comparisons for the, Endpoint vs. Tones stimuli, the Boundary vs. Endpoint stimuli and the Endpoint vs. Within Category stimuli.

Table 2: Areas of activation significant in planned comparisons, thresholded at voxel-level p<0.025, cluster-level p<0.01 (≥ 81 contiguous activated voxels). Coordinates indicate the maximum intensity voxel for that cluster, all coordinates are in Talairach and Tournoux space.

Anatomical Region	Maximum Intensity x	y	z	Number of activated voxels	Local Maximum (t value)
Boundary > Endpoint					
Left Inferior Frontal Gyrus	-50	26	24	259	3.581
Left Cingulate	-5	17	45	212	4.370
Endpoint > Within Category					
Left Anterior Cingulate	-2	47	3	156	3.541
Left Middle Temporal Gyrus	-44	-80	30	148	3.242
Left Precuneus	-2	-59	21	138	3.092
Boundary > Within Category					
Left Cingulate	-5	10	39	163	3.770
Left Subcortical	-17	8	6	136	3.402
Left Inferior Parietal Lobule	-62	-20	30	88	3.160
Endpoint > Tones					
Left Insula	-41	11	-7	274	2.490
Tones > Endpoint					
Right Insula	62	-14	12	608	3.031
Left Inferior Parietal Lobule	-65	-32	15	255	3.534
Left Cingulate	2	-20	45	215	5.543

Endpoint vs. Tones

The comparison of endpoint stimuli with tone stimuli yielded one cluster which was more active for endpoint stimuli than tones. This cluster was located in the anterior left insula, and extended deep into the lentiform nucleus, the caudate, and the putamen. In addition, three clusters were more active for tones than for endpoint stimuli. The largest of these (680 active voxels) was centered in the right insula, but extended largely along the right superior temporal gyrus and the right Heschl's gyrus, with some small extensions into the right inferior parietal lobule and precentral gyrus. A smaller cluster on the left side (255 active voxels) mirrored the larger cluster in placement, and was centered in the left inferior parietal lobule and left insula, with activation along the left Heschl's gyrus and left superior temporal gyrus. Finally, a cluster in the left and right cingulate was also more active for tones than endpoint stimuli.

Boundary vs. Endpoint

A comparison of endpoint stimuli and boundary stimuli yielded two significant clusters, both of which were more active for the boundary condition than the endpoint condition (Figure 5). The larger cluster was centered in the left inferior frontal gyrus, with activation extending dorsally into the middle frontal gyrus, and ventrally into the insula. The extracted hemodynamic time course for that cluster as shown in Figure 5 revealed modulation of activation, with boundary > within category > endpoint stimuli. This observation was confirmed in a statistical analysis of the time series data, using the mean cluster-wise percent signal change as the dependent variable, and time point and condition as the independent variables in a repeated measures, within-subject ANOVA. Data from time points at 2, 4 and 6 seconds were used in the analysis, as these times best capture the peak of the hemodynamic response in this area. This analysis showed a significant main effect of condition ($F(2,22)=13.305$; $p<0.001$). Post hoc Newman-Keuls tests revealed significant simple effects between all three stimulus types ($p<0.05$).

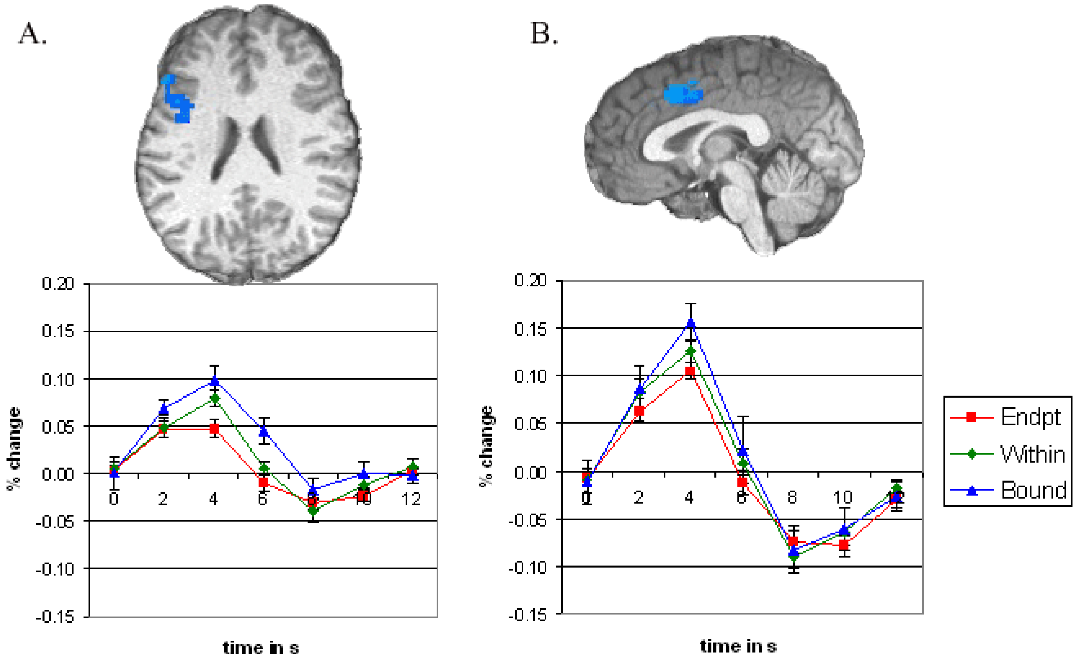

Figure 5. Clusters with significant activation differences in the Endpoint vs. Boundary comparison (p<0.01, corrected). Both clusters are more active for Boundary stimuli than Endpoint stimuli. A. shows the location of the left IFG cluster in an axial slice (z=24) and time series of activation for that cluster. B. shows the same for the cingulate cluster in a sagittal view (x=-2).

A second cluster centered in the left cingulate and extending into the right cingulate was also activated. As in the inferior frontal cluster, the time series data (see Figure 5) revealed a graded activation pattern, with boundary stimuli showing greater activation than within category, which in turn showed greater activation than endpoint stimuli. A repeated-measures ANOVA like the one reported above yielded a significant main effect of condition (F(2,22)=4.323; p<0.026). Post hoc tests revealed a significant difference between the endpoint and boundary conditions (p<0.05), but no other significant differences.

Endpoint vs. Within Category

Three clusters showed significant differences between the endpoint and within category stimuli, with more activation for the endpoint than within category stimuli in all clusters (see Figure 6). Activation emerged in a midline area centered in the left anterior cingulate, extending into the right anterior cingulate and the left medial frontal gyrus, in a posterior middle temporal area that extended into the angular gyrus, and in the precuneus bilaterally, with activation extending into the left and right posterior cingulate. However, an inspection of the time series data revealed that in each cluster, all conditions were deactivated, reaching peak deactivation at 6-8 seconds post-stimuli, and in each case, endpoint stimuli were less deactivated than within category stimuli (Figure 6).

Given that the left middle temporal gyrus and angular gyrus have both been implicated in speech processing, time series from this cluster was examined in more detail. Mean cluster-wise percent signal change values were submitted to an ANOVA as described above, the only difference being that data from time points at 4, 6, and 8 seconds were used. There was a main effect of condition ($F(2,22)=7.192$; $p<0.004$), and post hoc tests showed significant differences between the endpoint category and both the within category and boundary groups ($p<0.01$), but no difference between within category and boundary stimuli.

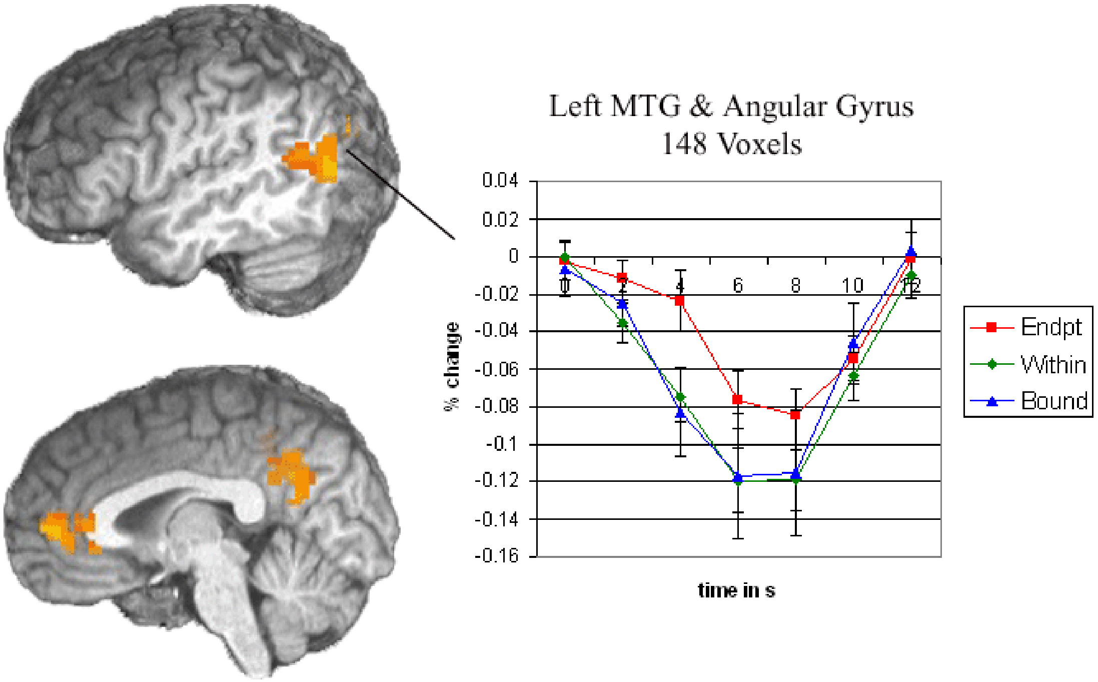

Figure 6. Clusters with significant activation differences for the Endpoint vs. Within Category comparison (p<0.01, corrected). All clusters showed deactivation for all stimulus types, with greater deactivation for Within Category stimuli than Endpoint stimuli. Sagittal views are shown at x= -50 (top) and x= -4 (bottom). The plot on the right shows the time series of activation extracted from the left MTG cluster.

2.3.2.3. ROI analysis

Under the hypothesis that auditory and auditory association areas such as Heschl's gyrus and the STG are sensitive to phonetic category structure, one would expect to find activation differences in these areas as a function of VOT category. No clusters centered in auditory areas were found in any of the VOT comparisons (although see the comparison between endpoints and tones, above). In order to further explore this hypothesis, a region of interest (ROI) analysis was carried out individually on the left and right STG and on Heschl's gyri bilaterally. Results showed no effect of VOT condition within left or right Heschl's gyrus (left: F(2,22)=0.891, right: F(2,22)=1.987). However, there was a significant main effect of VOT condition in the right STG (F(2,22)=5.779; p<0.01), and in the left STG, the effect approached significance (F(2,22)=3.235; p<0.059) (Figure 7). In the right STG, post hoc tests revealed that the main effect was driven by a difference in activation for boundary stimuli versus the endpoint and within VOT categories, (F>4.102; p<0.05 for both effects). The pattern of results was similar in the left STG, although the only comparison that reached significance was the boundary vs. within category comparison (F=3.026; p<0.05). The boundary vs. endpoint comparison approached, but did not reach significance (F=3.134), principally because in the Newman-Keuls test used, this comparison was subject to a higher statistical threshold than the boundary vs. within category comparison because the boundary vs. endpoint

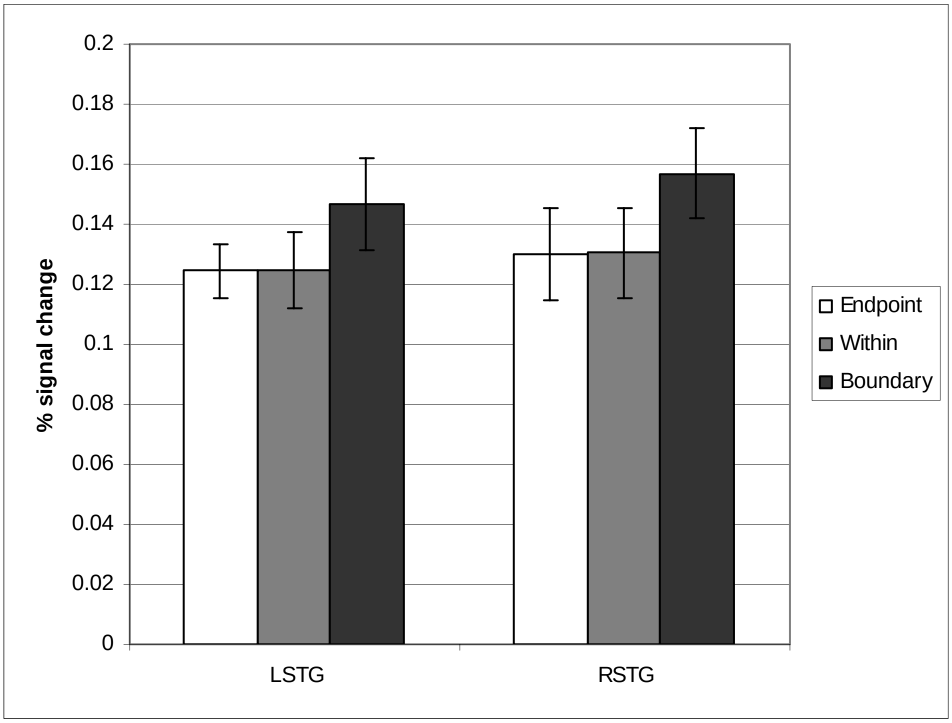

Figure 7. Mean percent signal change values for Endpoint, Within Category, and Boundary value stimuli from a region of interest analysis of the left STG and right STG. Error bars indicate standard error of the mean.

comparison was technically the largest comparison in the set.

A subdivision of the left STG into thirds along the anterior to posterior extent yielded the same general pattern of activation for all portions of the STG, with little difference in activation between endpoint and within category stimuli, and more activation for boundary stimuli (Figure 8). Only in the posterior STG region were there any significant differences between the speech stimuli (F(2,22)=3.790; p<0.039). Post hoc Newman-Keuls tests revealed no significant differences between endpoint, within category, or boundary stimuli within this area.

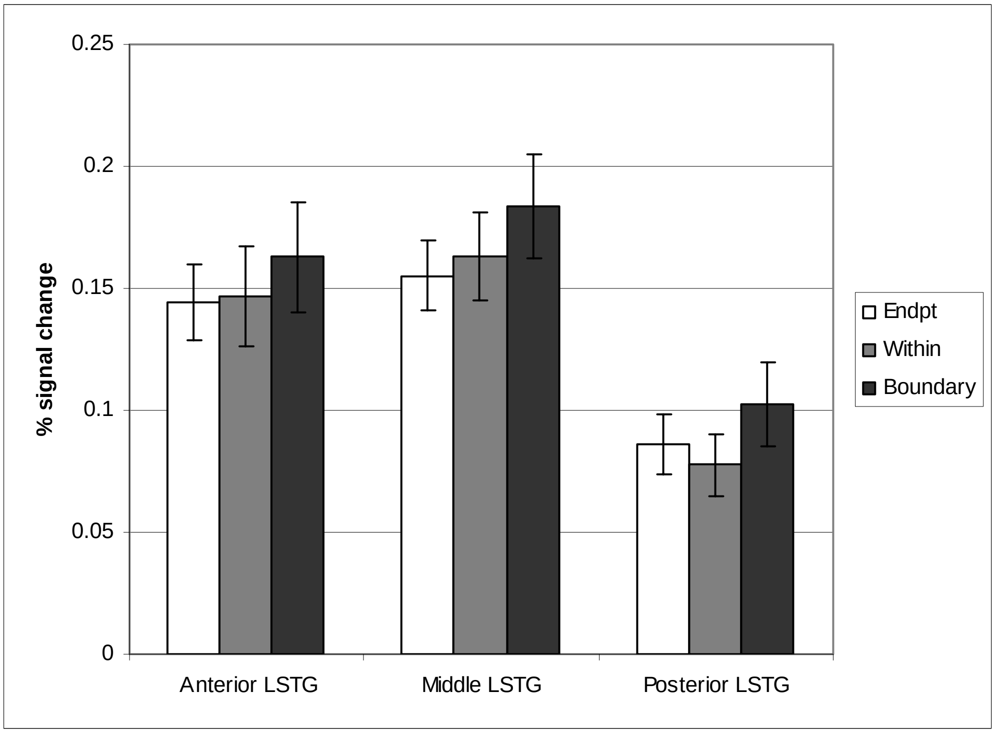

Figure 8. Mean percent signal change values for Endpoint, Within Category, and Boundary stimuli in the anterior, middle, and posterior portions of the left superior temporal gyrus. The anterior portion extended forward from a plane at y=-8, the middle portion extended between planes at y=-8 and y=-35, and the posterior portion extended posteriorly from a plane at y=-36. Error bars indicate standard error of the mean. The only area within which significant differences between stimulus types was observed was the posterior portion of left STG.

2.3.2.4. Regression analyses

2.3.2.4.1. Reaction time regression

Of interest is the degree to which the graded activation patterns in a number of areas correlates with the “difficulty” of the phonetic categorization task, as measured by reaction time. To this end, a regression was carried out which identified voxels in which a significant correlation with reaction time emerged that was not accounted for by subject or condition variables (see Methods).

A number of areas exhibited significant correlation between reaction time and percent signal change. These clusters are listed in Table 3. The largest of these was a bilateral posterior occipital cluster. This cluster is most likely an artifact, as it only appears at the edge of the scanning slab, and does not extend further into the slab. Additionally, both the left inferior frontal gyrus (Figure 9) and anterior cingulate showed a significant positive correlation between reaction time and percent signal change.

Table 3. Areas active in the reaction time regression analysis, thresholded at voxel-level p<0.025, cluster-level p<0.01 (≥ 81 contiguous activated voxels). Coordinates indicate the maximum intensity voxel for that cluster, all coordinates are in Talairach and Tournoux space.

Anatomical Region	*Maximum Intensity* x	y	z	*Number of Activated Voxels*	*Local Maximum (t value)*
Bilateral Posterior Occipital (edge of slab)	2	-95	12	1691	-3.880
L. Inferior Frontal Gyrus, L. Precentral Gyrus, L. Insula	-44	17	-1	150	3.423
R. Anterior Cingulate and corpus callossum	-2	29	9	124	-3.321
L. Lentiform nucleus	-16	-5	-7	115	-2.344
L. Caudate nucleus	-29	-26	-7	103	-5.291
R. Lentiform nucleus, R. Thalamus	-2	2	9	88	-2.686

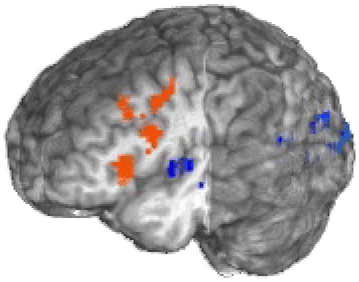

Figure 9. Results of the reaction time regression analysis. Rendered brain shown from the left hemisphere lateral aspect. Cutout extends anterior to y=-20 and left of x=-44.

2.3.2.4.2. Goodness of fit regression

The modulation of activation that emerged in a number of areas showing increased activation as stimuli approached the phonetic boundary raises the question of whether this modulation of activation reflects the phonetic categorization processes invoked by the task, or the increasing difficulty of the task as stimuli become more difficult to categorize. To investigate this issue, a regression analysis was carried out to decouple the effects of goodness of fit, which assumed a linear relationship between activation and proximity to the phonetic boundary, and difficulty of processing, which is assumed to be related to increases in reaction time as stimuli approached the phonetic boundary. To this end, we examined the portion of the variance in any voxel that could be absorbed by a “Goodness of Fit” (GF) regressor which was not already absorbed by linear, quadratic, and logarithmic reaction time regressors, (see Methods). Although GF and the linear reaction time regressor were significantly correlated, ($r(56)=0.2970$, $p<0.024$), when condition and subject variables are taken into account, this analysis was designed to factor out effects of stimulus difficulty on activation levels.

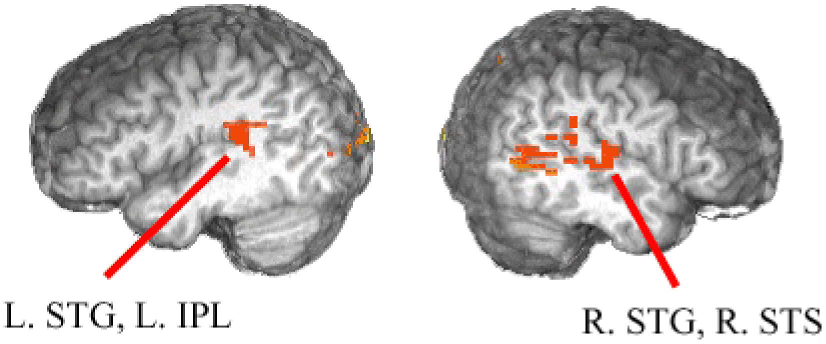

Figure 10. Results of the "Goodness of Fit" regression analysis. Sagittal views show bilateral STG clusters significant at a corrected p<0.01 (81 contiguous voxels, p<0.025). Left panel shows a rendered brain volume sliced at x=-44. Right panel shows a rendered volume sliced at x= 50.

Results revealed a number of areas in which there was a significant correlation between the signal and the GF regressor, over and beyond any correlation with reaction time (Table 4). Among these activations was a cluster in the left temporo-parietal junction. The maximum intensity voxel of this cluster fell in the left inferior parietal lobule, although the cluster fell primarily in the posterior left superior temporal gyrus, with some activation in the left middle temporal gyrus (Figure 10). A similar area in the right posterior superior temporal gyrus was also activated, which extended into the right middle temporal gyrus as well as the right Heschl's gyrus. This area was connected by a one-voxel bridge to the large bilateral occipital cluster noted in the cluster table, and as such is technically considered part of that cluster. Finally, a cluster in the right superior parietal lobule and right precuneus was also activated. Neither the cingulate nor the left inferior frontal gyrus showed any significant activation attributable to goodness of fit that was not already accounted for by the reaction time regressors.

Table 4. Areas active in the 'Goodness of Fit' analysis, thresholded at voxel-level p<0.025, cluster-level p<0.01 (≥ 81 contiguous activated voxels). Coordinates indicate the maximum intensity voxel for that cluster, all coordinates are in Talairach and Tournoux space.

Anatomical Region	*Maximum Intensity* x	y	z	*Number of activated voxels*	*Local Maximum (t value)*
Bilateral occipital, Right Superior Temporal Gyrus	20	-95	12	1023	3.849
Right Superior Parietal Lobule	17	-71	54	89	2.726
Left Inferior Parietal Lobule	-53	-38	26	82	3.840

2.4. Discussion

The results of this study show that there is a distributed neural network involved in the processing of phonetic category structure. These areas include the superior temporal gyrus bilaterally, the left inferior frontal gyrus, the left middle temporal gyrus extending to the angular gyrus, the left inferior parietal lobule, and the right and left cingulate. Different patterns of activation emerged in these areas as a function of the voice-onset time properties of the stimulus and its consequent phonetic category status as an endpoint, within category, or boundary stimulus, presumably reflecting different functional roles played by these areas in processing the phonetic categories of speech. In particular, frontal areas, including the inferior frontal gyrus and cingulate, showed modulation of activation as a function of graded category membership (boundary > within category > endpoint), and temporo-parietal areas, including the posterior MTG and angular gyrus, showed modulation of activation as a function of 'best fit' to the phonetic category (endpoint versus boundary and within). Of interest, these results emerged using a voice-onset time continuum that varied in 10 ms steps, indicating the sensitivity of the neural systems and the fMRI paradigm to acoustic fine structure along the phonetic processing stream.

2.4.1. Behavioral results

The behavioral results showed the typical categorical identification function for the perception of a voice-onset time continuum. Participants showed consistent identification of stimuli in the voiced and voiceless phonetic categories and chance performance to the boundary stimulus. The reaction-time data showed increased reaction-time latencies as a function of the ‘goodness’ of the stimuli as members of a phonetic category. The endpoint stimuli were responded to fastest, followed by the within category stimuli, and the slowest reaction-time latencies were to the boundary stimulus.

2.4.2. Functional results

2.4.2.1. Inferior frontal structures and graded category membership

Inferior frontal structures have been implicated in many aspects of language processing including phonetic processing (Burton, 2001). Additionally, these structures have been implicated in executive decisions with increased activation as those decisions become more difficult (Wagner, Pare-Blagoev, Clark, & Poldrack, 2001). In this study, the extent of phonetic processing necessary to map a token to a phonetic category largely covaried with the difficulty of the executive decision necessary to perform the task. That is, within category and boundary value stimuli are not only more difficult to map to a phonetic category and hence require increased processing to resolve phonetic category membership, but the increased reaction-time latencies as the stimuli approach the phonetic boundary indicate that the executive decision becomes more difficult as well. As such, either the hypothesis that frontal structures have a role in phonetic processing or the hypothesis that frontal structures have a role in executive decisions would predict that modulation of activation should emerge as a function of the stimulus category with the greatest activation for boundary value stimuli, less activation for within phonetic category stimuli, and the least activation for the endpoint stimuli. The results of the analysis comparing the endpoint and boundary stimuli are consistent with this prediction. There was greater activation for boundary stimuli versus endpoint stimuli in the left IFG extending dorsally into the MFG and ventrally into the insula and, as well, for the left cingulate extending into the right cingulate. The extracted hemodynamic time course for these clusters showed graded activation for phonetic category membership, with boundary > within category > endpoint stimuli.

The question remains whether this pattern of graded activation reflects properties of phonetic category structure or reflects the difficulty of the decision. In a study aimed at distinguishing sensory and decision processes in phonetic perception, Binder and colleagues (2004) found that activation in a portion of the bilateral opercula and insulae correlated with reaction time but not with accuracy. They propose that accuracy reflects sensory components and reaction time reflects decision components, and hence attributed an executive role to medial portions of the operculum. What is not clear is whether the left frontal activation in the current study is due to stimulus difficulty, to phonetic processing, or to some combination of the two. The regression analysis provided a possible means of distinguishing these possibilities.

In particular, the results of the regression analyses showed no significant activation in the left IFG over and above that accounted for by reaction time, suggesting that activation in the left IFG is closely related to the stimulus difficulty and presumably the consequent response conflict (Gehring & Knight, 2000; MacDonald, Cohen, Stenger, & Carter, 2000). Likewise, all of the cingulate activation could be accounted for solely by reaction time. In contrast, significant activation was revealed in the bilateral posterior superior temporal gyri, which was correlated with the ‘goodness of fit’ of the stimulus to its phonetic category, but not with difficulty. This result supports the view that both the cingulate and the left IFG have a role in executive processes, while the sensory component of phonetic processing is subserved by the bilateral posterior superior temporal gyri.

There is another aspect to the processing of phonetic category structure which could account for the modulation of activation observed in the left IFG. Not only are the non-endpoint tokens more difficult to map onto a phonetic category, but they also are processed under conditions of increased phonetic competition. A within category stimulus is closer in acoustic space to the contrasting phonetic category, and a boundary value stimulus is equivocal in its category membership. Thus, the increased frontal activation could also reflect phonetic competition (Thompson-Schill et al., 1997). Such increases in phonetic competition would result in increases in decision difficulty, and hence reaction time, and would therefore be accounted for in the reaction time regression analysis.

2.4.2.2. The cingulate and response competition

In the present study, graded activation was seen in the cingulate, with more activation the boundary stimulus, less for within-category stimuli, and least for endpoint stimuli. Like inferior frontal structures, the cingulate gyrus has been implicated in aspects of cognitive control. The cingulate is particularly implicated when there is competition at the level of the response (Milham et al., 2001; van Veen, Cohen, Botvinick, Stenger, & Carter, 2001). As such, the graded pattern of activation observed in the cingulate may reflect competition between the /da/ and /ta/ responses, rather than competition at the phonetic level of processing.

2.4.2.3. Tempo-parietal areas and best fit to the phonetic category

Several significant clusters, among them one in the posterior tip of the left MTG extending into the angular gyrus (AG), emerged showing significantly greater activation for endpoint stimuli versus within category stimuli. Thus, in contrast to the modulation of activation described above for the other clusters, these clusters appeared to show the *greatest* activation for the stimulus on the continuum that was the 'best fit' to its phonetic category. However, time series analysis of the MTG-angular gyrus cluster showed deactivation for all stimuli, with the greatest deactivation for both the within category and boundary value stimuli compared to the endpoint stimuli. Thus, this area shows differential sensitivity to the 'best fit' stimuli versus all other speech stimuli, with activation levels for all stimuli remaining below baseline.

Deactivation in parietal areas including the angular gyrus has been shown in a number of studies. There has been considerable discussion about the role that deactivation plays. Some have proposed that it reflects reallocation of processing resources from deactivated areas to those areas involved in the task (McKiernan, Kaufman, Kucera-Thompson, & Binder, 2003), with the extent of deactivation a function of the processing resources required to do the task. Other have suggested that deactivation reflects a 'gating' role, i.e. controlling input to areas not involved in the task (Gusnard & Raichle, 2001; Shulman et al., 1997). Binder et al. (1999) suggest that these areas are involved in ongoing processes during resting states, processes which involve access to conceptual knowledge, and hence these areas become deactivated when exogenously presented input requiring attention and response are presented.

Although the proposals about the role of deactivation vary, they all share the view that deactivation indicates that these areas are not actively involved in the processing of the stimuli in the task. However, the results of the current study suggest that indeed the MTG/AG are actively involved in phonetic processing and appear to be ‘tuned’ to the best match of the acoustic input to the phonetic category or to the label for that category. If this area were to reflect either reallocation of resources or ‘gating’ input with the extent of activation a function of the degree of resources required, then graded activation should have emerged for the stimulus categories, as was found in the IFG and the cingulate. However, the activation patterns in the MTG/AG showed no differences in the patterns of deactivation for the poorer exemplars (the within category and boundary value stimuli) (see Figure 6), and significantly less deactivation for the best exemplars (the endpoint stimuli). Thus, the MTG/AG did not show increased deactivation as a function of stimulus difficulty, but rather showed sensitivity only to the ‘best’ fit of the phonetic category, suggesting that this area is actively involved in processing the stimuli.

As discussed previously, left temporo-parietal areas including the angular gyrus have also been implicated in aspects of phonetic-phonological processing, including the manipulation and storage of information in phonological store (Paulesu et al., 1993), and access to the rules of one’s phonological grammar (Jacquemot et al., 2003). Additionally, evidence from the aphasia literature has linked lesions in this area to deficits in the discrimination and identification of phonemes (Caplan et al., 1995). Recent neuroimaging results showed increased activation in the left lateral temporo-parietal region under conditions of increased lexical competition, consistent with the view that this region is involved with mapping sound structure to lexical form and accessing the sound shape of the word from the phonological store (Prabhakaran et al., submitted). Taken together, the results suggest that the MTG/AG is involved in the mapping of the acoustic input to the representation of the phonetic category. Deactivation occurred for all stimulus types, with the least deactivation for the stimulus which is the best exemplar of the phonetic category. Unlike frontal areas, the pattern of results observed in this area was not graded, that is, there was a clear distinction between endpoint stimuli and all other syllable stimuli. While sensitivity to within-category phonetic structure was observed (i.e. endpoint stimuli are different from within-category stimuli), there was no difference between boundary and within-category stimuli. This suggests that the sensitivity of this area is not to phonological categories per se, but to the tokens which are the best representations of those categories.

Evidence from lesion studies suggests that the inferior parietal area including the angular gyrus is actively involved in lexical access or naming. Classic pure anomic patients have damage to the left temporo-parietal junction including the angular gyrus and show selective deficits in naming (Geschwind, 1965; Goodglass, 1993). These patients understand words and show intact conceptual knowledge. However, they appear to be unable to access the sound shape or ‘name’ of the word. A recent study showed that lesions in the left middle temporal gyrus correlate strongly with deficits in comprehension on a single word level (Dronkers, Wilkins, Van Valin, Redfern, & Jaeger, 2004). The results are also compatible with the hypothesis that the temporo-parietal junction is involved in access to lexical representations. The phonetic categorization task is essentially a naming task. Thus, it is also possible that this area is recruited for implicit naming of the stimulus as a ‘d’ or a ‘t’, with the least deactivation for the stimulus that is the best exemplar of the category.

2.4.2.4. Role of the superior temporal gyrus

Although the mean activation maps showed extensive activation in the superior temporal gyrus bilaterally, analyses comparing the endpoint, within category, and boundary value stimuli failed to reveal any areas of significant difference in either the left and right superior temporal gyri. These results suggest that the STG is involved in the early analysis of the acoustic properties of speech, but sends this information to other areas for higher levels of processing relating to phonetic category structure. An ROI analysis investigating the percent signal change across the VOT categories in Heschl’s gyrus showed no modulation of activation as a function of VOT. Thus, it appears that Heschl’s gyrus is involved in the early acoustic analysis of the sensory information inherent in the stimuli. In contrast, the ROI analysis revealed greater activation for the boundary value stimuli compared to the within category and endpoint for both the right and left STG indicating sensitivity to VOT differences. Consistent with these results, the regression analysis showed that the bilateral posterior superior temporal gyri were sensitive to the ‘goodness of fit’ of a stimulus to its phonetic category, but were insensitive to the difficulty of the stimulus, as measured by reaction time. This evidence further supports a role for the bilateral superior temporal gyri, particularly the posterior portions, in the early phonetic processing of speech stimuli. That modulation of activation emerged between the boundary stimuli versus the within category and endpoint stimuli in this analysis is consistent with electrophysiological findings using the MMN paradigm. As described earlier, Sharma and Dorman (1999) showed a larger MMN for between phonetic category stimulus pairs than for within phonetic category stimulus pairs in a voice-onset time continuum. Nonetheless, although these results support the hypothesis that the STG bilaterally is sensitive to phonetic category structure, such a conclusion needs to be made with caution. Although MMN sensitivity has been observed for VOT and also for a number of vowel contrasts, it has not emerged consistently for another phonetic contrast, place of articulation in stop consonants (Phillips, 2001). Thus, it is not clear whether the modulation of the MMN reflects phonetic category structure per se or natural acoustic discontinuities in these continua.

Although the superior temporal gyri do seem to show sensitivity to differences in VOT, it is of interest that no difference was found in activation levels for tokens within a phonetic category. That is, within-category and endpoint stimuli showed approximately the same level of activation. This contrasts with areas such as the left inferior frontal gyrus and the cingulate, which do show activation differences between endpoint and within-category stimuli. Given that these frontal areas, which are further down the processing stream than the superior temporal lobes, show such graded activation patterns, it must be the case the superior temporal gyri are also sensitive to such differences. There are several reasons why activation differences between endpoint and within-category stimuli might not emerge in the superior temporal gyrus. First, it may be the case that, although there is differentiation in the neural code between members of a phonetic category, as is suggested by the N100 results from the ERP/MEG literature, this differentiation does not translate into a difference in blood oxygenation level in this area.

Second, this lack of difference may arise due to the structure of the phonetic category itself. It has been suggested that the structure of acoustic-phonetic space within a phonetic category is not linear in nature. Kuhl and colleagues (see Kuhl, 1991) have argued that perceptual space is warped within a phonetic category such that there is greater sensitivity to acoustic-phonetic differences near the phonetic category boundary than near the center of the phonetic category. This may account for the increased activation for a stimulus which falls on the phonetic boundary, while activation levels for tokens nearer to the center of the phonetic category are relatively similar.

Of interest is the failure in this study to find differences among syllable stimuli in the left anterior temporal lobe. The left temporal lobe lateral and anterior to Heschl's gyrus has been implicated in analysis of spectrotemporal cues in speech, with eventual mapping of these cues to the lexicon (Scott & Johnsrude, 2003), and activation in a similar area bilaterally has been shown to correlate with sound identification accuracy (Binder et al., 2004). An informal post-hoc analysis of activation in the left STG showed that although anterior, middle, and posterior thirds of the left superior temporal gyrus showed similar patterns of activation, with boundary stimuli more active than either endpoint or within category stimuli, this pattern only reached significance in posterior portions of the left STG. One possible explanation for the lack of a significant difference in anterior areas is that the greatest signal dropout occurred in the lateral anterior temporal lobe, and thus differences did not emerge in this area. A second possibility is that the arbitrarily-defined thirds of the STG include more than one functionally distinct area. Indeed, some have suggested that a gradient of functionality exists along the medial to lateral surface of the superior temporal plane as well as in the anterior to posterior direction (Scott &

Johnsrude, 2003). A third possibility is that although the anterior temporal lobes may be involved in spectro-temporal analysis (Scott & Johnsrude, 2003; Wise et al., 2001), the differences in the spectro-temporal properties of speech stimuli in this study are so minimal so as to engage the analysis process equally. All VOT stimuli in this study shared the same spectral properties, and the variation in the temporal property of VOT, although phonetically meaningful, is on the order of tens of milliseconds. The fact that posterior STG shows significantly different activation for boundary stimuli versus other speech stimuli, and also shows activation patterns sensitive to the 'goodness of fit' of stimuli, taken together with the observation that an adjacent area in the left posterior MTG shows preferential sensitivity to the stimuli which are the 'best fit' to a phonetic category, suggests that a posterior stream of phonetic processing may be engaged in mapping acoustic input onto the abstract representation which constitutes the phonetic category.

2.5. Conclusion: Role of phonetic category structure in language processing

The results of this study suggest that the neural system is not only sensitive to phonetic category structure but the details of phonetic category membership are also retained throughout the phonetic processing stream. Even in the MTG/AG, which showed a binary, 'categorical,' response function, activation differences were observed between stimuli within a phonetic category, and sensitivity to within-category structure was preserved. One issue is whether the sensitivity to this information is a function of the demands of the phonetic categorization task itself or whether sensitivity to phonetic category structure has consequences at higher levels of processing. Evidence from behavioral studies suggests that phonetic category structure influences not only phonetic categorization processes but also influences lexical access and ultimately access to the lexical-semantic network (Andruski et al., 1994; McMurray et al., 2002; McMurray et al., 2003; Utman et al., 2001). In particular, McMurray and colleagues have shown that eye movements to picturable nouns which are distinguished only by the voicing of the initial consonant (i.e. ‘bear’ and ‘pear) are affected by within-category reductions in VOT. Additionally, the magnitude of semantic priming, e.g. *c*at- dog* shows less semantic priming than *cat-dog* (Andruski et al., 1994). These results suggest that phonetic category structure plays a role in higher levels of language processing, and would presumably influence the neural activation patterns during such processing. Further research is needed to investigate this issue. However, taken together with the results of the current study, it suggests that sensory information is retained and used beyond the sensory system that is responsible for its analysis.

As yet unresolved is the issue of whether patterns of graded activation in frontal areas reflect processing required to map a stimulus onto a phonetic category which varies as a function of that stimulus’ goodness of fit, or whether such a pattern instead reflects competition between the two phonetic categories. While the regression analyses suggest that activation in frontal areas largely covaries with the difficulty of the task, processing as a function of goodness of fit and competition might both be expected to contribute to such difficulty, and as such are conflated in this design. The goal of the next experiment is to decouple goodness of fit and competition by comparing activation in response to poor tokens which do not fall near the phonetic category boundary to those that do fall near the boundary. This design likewise allows us to test the hypothesis generated by this experiment that superior temporal areas are sensitive to the goodness of fit of tokens to a phonetic category, while remaining relatively insensitive to the effects of competition.

3. CHAPTER 3: EXPERIMENT 2, COMPETITION VS. GOODNESS OF FIT

3.1. Introduction

As discussed in Chapter 2, a number of properties of the cognitive architecture of the system influence the phonetic categorization task. In order to perform such a task, listeners must first analyze the basic acoustic-phonetic properties of the stimulus. Those properties then map on to a matching phonetic category. In a model of language processing such as TRACE (McClelland & Elman, 1986), the phonetic categories of speech form a network of connections such that the activation of one phonetic category may compete with or influence the activation of another phonetic category. In such a model, a poorer exemplar of the phonetic category activates its category representation more weakly than a good exemplar. This weak activation results in greater time for the activated phonetic category to reach a decision threshold, and should therefore increase the difficulty of the phonetic categorization decision. Thus the ease with which the system resolves on a phonetic category depends on the token's goodness of fit to the phonetic category. In addition, at this stage of processing there may be competition between the activated phonetic category and other similar categories. Although any phonetic category may in theory compete with any other phonetic category, more competition is expected when the input is more similar to other categories in the language inventory—for instance if the stimulus falls near a phonetic category boundary. Because the phonetic categorization task requires an explicit response, an appropriate response must be selected from the response set, and that response must be mapped onto a motor program, which ultimately results in the button press. Thus, the goodness of fit of a phoneme to its phonetic category as well as the degree of competition a phoneme is expected to undergo as a function of its similarity to other phonemes are both intrinsic parts of the phoneme categorization process.

As yet, the neural substrates responsive to phonetic competition and goodness of fit are unclear. In Experiment 1, greater activation was observed for stimuli which fell on the phonetic category boundary than to other stimuli. This activation emerged in both posterior (STG) as well as anterior (cingulate, L IFG) structures. A regression analysis in that study suggested that activation in frontal areas was closely linked to difficulty of the phonetic decision, whereas activation in posterior regions was linked to the goodness of fit of the stimuli. It was proposed that the increase in activation in the left IFG was linked to its hypothesized role in resolving competition (c.f. Thompson-Schill et al., 1997), as those tokens nearest to the category boundary should be subject to the largest amount of competition from the contrasting phonetic category. However, in this study, the degree of competition and the ‘poorness’ of fit to a phonetic category were confounded. Thus, it is as yet unclear whether the regression analysis reported in Experiment 1 was able to truly decouple the relative effects of competition and goodness of fit.

The goal of the present study is to attempt to decouple competition from goodness of fit and to investigate the relative sensitivity of anterior and posterior brain structures to these two aspects of phonetic processing. To this end, we will examine activation patterns for stimuli which are acoustically very far from the phonetic boundary. Although such stimuli are poorer exemplars of their phonetic category, they are minimally influenced by competition from the contrasting phonetic category. Thus, it may be possible to separately assess the effects of competition and goodness of fit.

Internal structure of voiceless and voiced stop categories

As demonstrated in Experiment 1, structure exists within a phonetic category, with some members of that category serving as better exemplars of that category than others. This is apparent despite the fact that subjects are often poor at discriminating poor and good exemplars within the same category. For instance, subjects show increased reaction times to stimuli along a VOT continuum as they approach the category boundary (Experiment 1). It is also the case that stimuli which are not equivocal in their phonetic category membership may also be poorer members of a phonetic category. Volaitis and Miller showed that voiceless tokens with very long VOTs of 165 msec and greater were judged as poorer phonetic category members than those with shorter VOTs (Miller & Volaitis, 1989; Volaitis & Miller, 1992). Kessinger and Blumstein (manuscript) also showed that stimuli beginning with voiceless stops which were increased by 4/3rds (mean VOT=180 msec, range: 136-225 msec) were rated as poorer exemplars than either unaltered stimuli or those that were increased by 2/3rds (mean VOT=129 msec, range: 103-154 msec). No difference was found between unaltered stimuli and those increased by 2/3rds. It appears, therefore, that phonetic category goodness for voiceless stops begins to decrease for tokens with VOTs of approximately 165-180 msec or greater.

Less is known about the structure of the voiced end of the phonetic category in English speakers. Voiced stops may have VOTs from 0 to about 30 msec, but many speakers of English also produce prevoiced stops, where the onset of voicing precedes the burst. Because VOT is defined as the latency between the burst and the onset of voicing, prevoiced stops have negative VOT values. A study by Ryalls et al. (Ryalls, Zipprer, & Baldauff, 1997) which investigated VOT production as a function of race and gender found that the mean VOT for voiced stops ranged from -46 msec to +12 msec VOT depending place of articulation of the stop consonant, race and gender. More specifically, VOTs for /d/ had the most prevoicing (average VOT ranged from -46 to -7 msec depending on the race and gender of the speaker). Thus, it is likely that tokens with prevoicing of at least -46 msec are still good exemplars of the /d/ category. To date, no studies have investigated the perceptual effects of extreme amounts of prevoicing. It is hypothesized that stops with very long prevoicing will, like those with very long VOTs, be perceived as poorer members of their phonetic category, and hence will show slower reaction times in identification tasks.

Both phonetic competition and goodness of fit are expected to vary as a function of the internal category structure of a phoneme. Competition is expected to increase as stimuli become more similar to the contrasting phonetic category, and as such, competition should be maximal near the category boundary. Goodness of fit, in contrast, is expected to be best for tokens near the center of the phonetic category, and worse for tokens further from that center, be they near the boundary, or on the ends of the continuum (see Figure 11 for a schematic of the hypothesized effects of competition and goodness of fit along a VOT continuum).   In this study, as in Experiment 1, subjects performed a phonetic categorization task (i.e. “/da/” or “/ta/”) on stimuli from a VOT continuum. Activation patterns for tokens both near the category boundary (“Near-Boundary”) and at the periphery of the continuum (“Extreme”) were contrasted with ‘good’ tokens (“Exemplar”) in order to isolate the effects of goodness of fit. Likewise, activation patterns to tokens near the phonetic boundary (Near-Boundary) were contrasted with tokens further away from the phonetic boundary (Extreme and Exemplar),

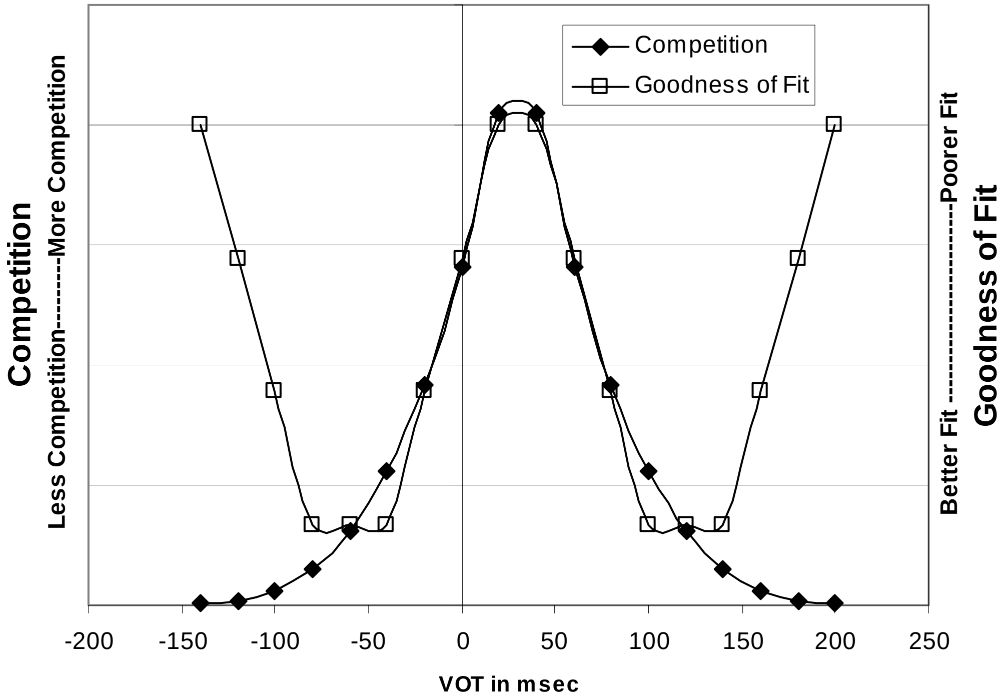

Figure 11. Hypothesized pattern of competition and goodness of fit along a voicing continuum. Y-axis scale is in arbitrary units

in order to isolate the effects of competition on activation levels. It is hypothesized that greater activation should be seen in frontal areas for Near-Boundary stimuli due to the great competition these tokens undergo. In contrast, activation for Near-Boundary and Extreme tokens should be greater than that for Exemplar tokens in the STG if this area is, as hypothesized, sensitive to goodness of fit.

3.2. General Methods: Stimulus Generation

Stimuli were generated using the KlattWorks interface (McMurray, in preparation) to the Klatt synthesizer, which combines cascade and parallel methods of synthesis (Klatt, 1980). First, a four-formant pattern with a VOT of 0 msec was generated. Formant values were patterned after Stevens & Blumstein (1978) values given for /da/. Onset values for F1, F2 and F3 were 200, 1700, and 3300 Hz. There was a piece-wise linear transition of 40 ms to steady state values of 720, 1240, and 2500 Hz. F4 remained steady at 3300 Hz throughout the stimulus. An initial burst was synthesized by manipulating the amplitude of frication, and aspiration noise was added by manipulating the amplitude of aspiration. For the 0 msec token, the aspiration noise was present during the burst alone. The amplitude of voicing started at the onset of the stimulus at a value of 57 dB, transitioned to 60 dB over the next 5 msec, and remained at 60 dB until 30 msec from the end of the stimulus, at which point it dropped piece-wise linearly to 30 dB over those 30 msec. Stimuli with increasing VOTs were generated from this token by delaying the onset of voicing and increasing the duration of the aspiration noise in 10 msec increments. In this way, stimuli were generated varying in 10 msec steps from 0 msec to 190 msec VOT. Total stimulus duration was kept constant at 330 msec by removing 10 msec increments from the steady state portion of the middle of the vowel.

Prevoiced stimuli were also generated from the 0 msec VOT token. The amplitude of voicing for the parallel synthesizer was increased to 50 dB prior to the burst in successive 10 msec intervals. At the same time, a portion of the midpoint of the vowel which was of the same duration as the added prevoicing was removed in order to preserve the stimulus length at 330 msec. Prevoiced stimuli ranging from 0 msec to -190 VOT were generated using this method.

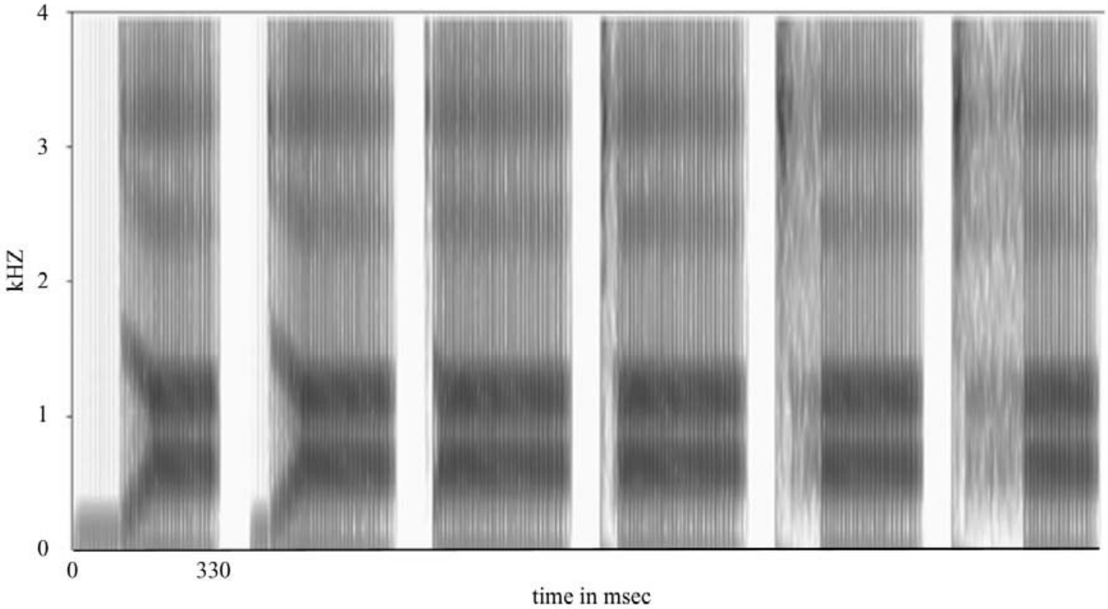

Figure 12. Spectrogram of stimuli used in Experiment 2 for Pilot 3 and fMRI experiment. From left, stimuli have VOTs of -100, -40, 20, 40, 100, and 160 msec.

For the stimulus with VOT of 0 msec, F0 was 104 Hz at the onset of the stimulus (coincident with the burst), increased to 125 Hz at 75 msec after the burst, then fell linearly to 97 Hz at the end of the stimulus. For stimuli with a VOT greater than or equal to 0 msec, this pitch contour was kept constant. For prevoiced stimuli, F0 was 104 during the prevoicing period, then followed the same contour as for stimuli with VOTs greater than 0. Spectrograms of representative stimuli are shown in Figure 12.

3.2.1. Pilot 1: Full Continuum

A first pilot experiment was run in order to ensure that subjects heard stimuli as exemplars of either the /da/ or /ta/ category, to determine where the phonetic category boundary fell in this stimulus set, and to ensure that subjects perceived stimuli categorically. On the basis of previous research, it was expected that subjects would show significantly longer reaction times for stimuli near the phonetic category boundary (see Experiment 1), and also for stimuli with extremely long VOTs (Kessinger & Blumstein, manuscript). It was predicted that a similar increase in RT would occur for tokens with extreme prevoicing (negative VOTs), as they should be poorer exemplars of the voiced category.

Subjects heard tokens from the -190 to 190 msec VOT continuum. The stimulus set included stimuli in 20 msec VOT steps from -190 to -10 msec VOT and from 70 msec to 190 msec VOT. Stimuli were included at 10 msec steps from -10 msec to 70 msec in order to more finely sample the region in which the phonetic category boundary was expected to fall. There were a grand total of 24 stimuli varying in VOT. Subjects heard 15 repetitions of each stimulus, and stimuli were evenly divided into three separate runs.

Eight pilot subjects participated in this experiment for modest monetary compensation. Subjects were asked to listen to each stimulus and to decide whether it began with a /d/ or /t/ sound, and to press a corresponding button as quickly and accurately as possible.

Reaction time and percent '/da/' responses for the eight pilot subjects are displayed in Figure 13. Subjects had no trouble perceiving stimuli as /ta/ and /da/, and categorization data revealed that subjects responded with near-ceiling consistency for all stimuli except those which fell at 20 and 30 msec VOT, which is the region in which the phonetic category boundary typically falls (see Experiment 1). Reaction time data revealed a marked peak in RTs at 20-30 msec VOT, and another, smaller peak at -10 msec VOT. Reaction times increased slightly at the extreme prevoiced (-190 msec VOT)

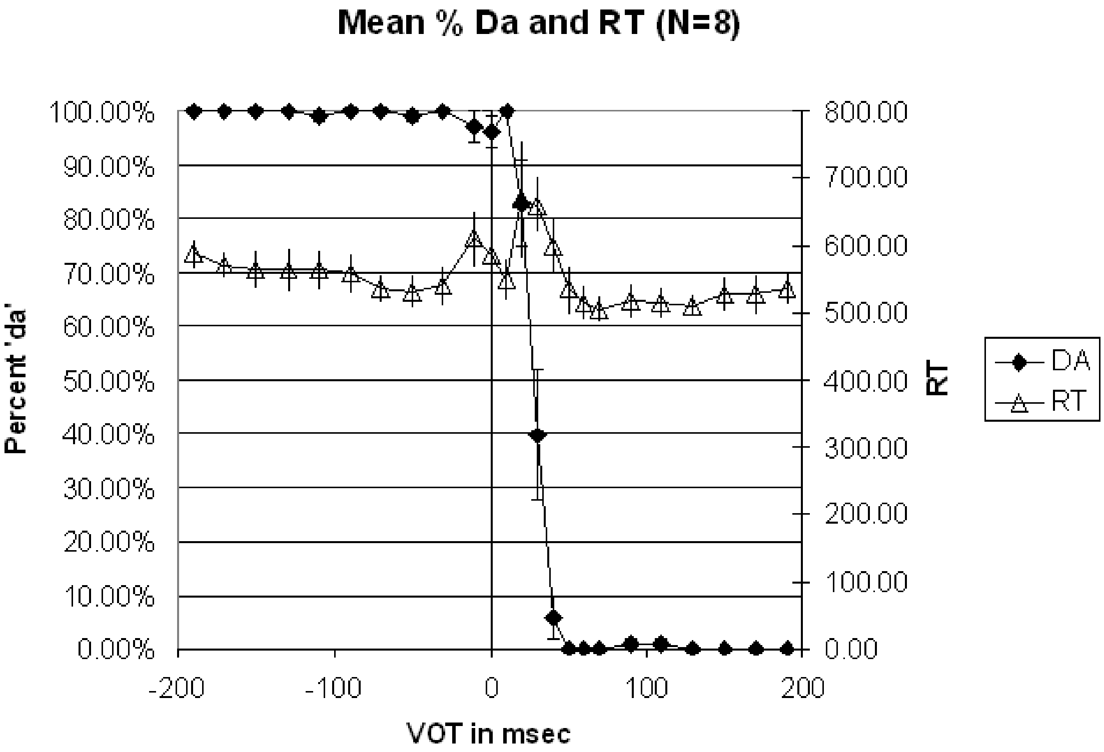

Figure 13. Reaction time and identification data from Pilot 1 (n=8).

and extreme lengthened (190 msec VOT) ends of the continuum, but not as markedly as expected.

It is possible that the predicted effect of increased RTs to extremely increased VOTs might be attenuated in the larger continuum because of the presence of many other similarly anomalous exemplars in the stimulus set. To this end, a second pilot was carried out which included two exemplars from close to the phonetic category boundary, two exemplars near the centers of the /da/ and /ta/ phonetic categories, and two exemplars from opposite ends of the continuum. These stimuli were chosen with the goal of equating reaction times for stimuli near the category boundary and those near the extreme ends of the continuum, while keeping identification performance for all stimuli consistent.

3.2.2. Pilot 2: Partial Continuum

Stimuli were chosen on the basis of Pilot 1 to contain one stimulus from each of the /da/ and /ta/ categories which was from the extreme end of the continuum (Extreme), one which fell near the center of the phonetic category (Exemplar), and one which fell near the phonetic category boundary (Near-Boundary). Stimuli were also chosen with the goal of having relatively even steps between points on the VOT continuum. On this basis, two Extreme stimuli were chosen (-110 and 160 msec VOT), two Exemplar stimuli were chosen (-50 and 100 msec VOT), and two Near-Boundary stimuli were chosen (10 and 40 msec VOT). All stimuli were 60 msec distant from their nearest neighbors, except those near the phonetic category boundary, which were 30 msec distant from each other. Subjects heard a total of 20 repetitions of each stimulus

Ten pilot subjects participated in this experiment for modest monetary compensation. As in Pilot 1, subjects were asked to listen to each stimulus and to decide whether it began with a /d/ or /t/ sound, and to press a corresponding button as quickly and accurately as possible.

Reaction time means for the ten pilot subjects, grouped by stimulus category (Extreme, Exemplar, Near-Boundary) are displayed in Table 5. Subjects were extremely consistent in the identification of all stimuli, with all stimuli categorized consistently 99% of the time or better. Subjects showed slowing of reaction times for stimuli at the extreme ends of the continuum (-110 and 160 msec VOT), but little slowing for those near the phonetic category boundary (10 and 40 msec VOT). This observation was confirmed by a within-subjects repeated measures ANOVA on reaction times for the Extreme, Exemplar, and Near-Boundary categories of stimuli. This ANOVA revealed a significant effect of VOT category ($F(2,18)=3.975$, $MSE=782.3$, $p<0.037$). Post-hoc Newman-

Keuls tests revealed significant differences in reaction time between the Extreme condition and the other two conditions, but no difference between the Near-Boundary and Exemplar conditions. Inspection of the RT data for individual stimuli suggested that this might be due to the decreased reaction time shown to the 10 msec VOT token, which was also observed in Pilot 1. A third pilot was carried out in which the 10 msec stimulus was replaced with a 20 msec token, with the goal of equating reaction times for Near-Boundary and Extreme conditions.

Table 5. Reaction time means for Pilot 1 (n=10). Standard deviation indicated in parentheses.

Stimulus	*Mean RT*
Extreme (-110 and 160 msec VOT)	593.9 (104.5)
Exemplar (-50 and 100 msec VOT)	560.5 (98.4)
Near-Boundary (10 and 40 msec VOT)	567.4 (108.4)

3.2.3. Pilot 3: fMRI Set

Stimuli were chosen for this experiment on the basis of the results obtained in Pilots 1 and 2. The 10 msec token from Pilot 1 was replaced with at 20 msec token, and stimuli on the /da/ end of the continuum were increased by 10 msec VOT in order to preserve the even steps between stimuli. Thus, the stimulus set consisted of two Extreme stimuli (-100 and 160 msec VOT), two Exemplar stimuli (-40 and 100 msec VOT) and two Near-Boundary stimuli (20 and 40 msec VOT). As in Pilot 2, subjects heard 20 repetitions of each stimulus.

Nine subjects participated in this experiment for modest monetary compensation. The subject instructions were the same as for Pilots 1 and 2. Reaction time and percent /d/ responses are displayed in Figure 14A. The same reaction time data is grouped by stimulus category in Figure 14B. Subjects again showed near-perfect consistency in identification of all stimuli, with categorizations of 96% consistent or better. Reaction times showed slowing both for Extreme stimuli at the continuum periphery as well as

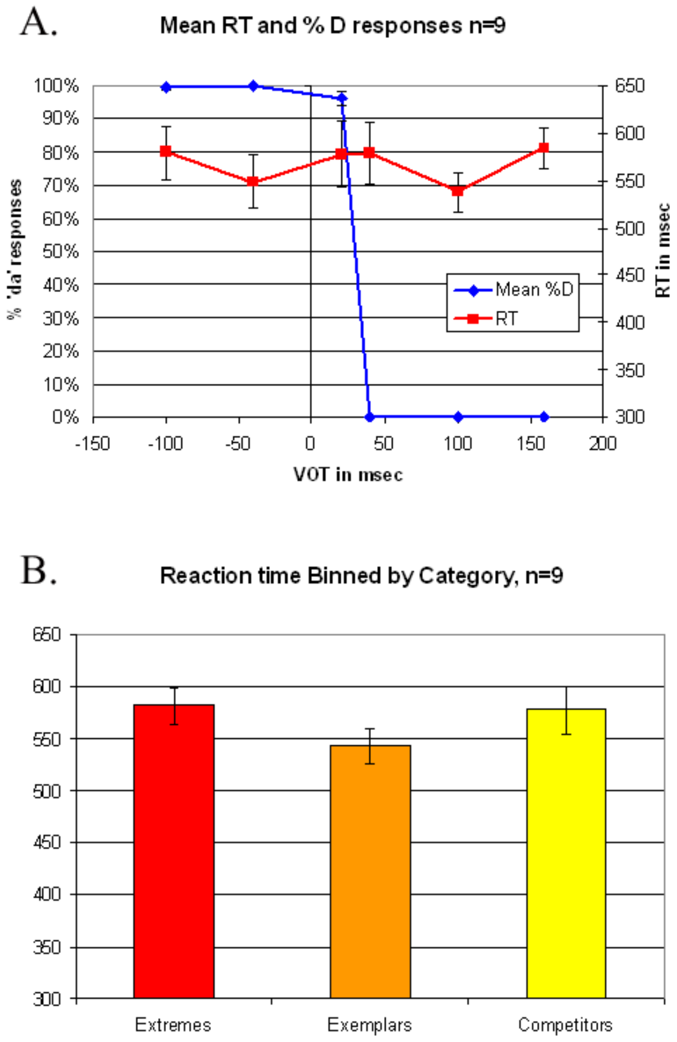

Figure 14.   Reaction time and identification results from Pilot 3 (n=9). Upper portion (A) shows results for individual tokens along the VOT continuum. Bottom (B) shows the same reaction time results, binned by stimulus category.

Near-Boundary stimuli. This observation was confirmed by an within-subject repeated measures ANOVA on the mean RTs for the Extreme, Exemplar, and Near-Boundary categories, which showed a main effect of VOT category on RT (F(2,16)=3.087, MSE=1304.7, p<0.073) which approached significance. The magnitude of difference between RT categories was actually larger in this pilot experiment than in Pilot 2, and standard deviations were very similar. The magnitude of difference between Extreme and Near-Boundary RT measures was minimal (4 msec). This stimulus set was therefore selected for use in the fMRI experiment.

3.3. Experiment 2: Goodness of Fit vs. Competition

3.3.1. Introduction

This fMRI study was designed to investigate the sensitivity of anterior and posterior neural structures to the goodness of fit of a token to its phonetic category versus the competition which arises when stimuli fall near a category boundary. This was investigated using a phonetic categorization task on stimuli which ranged from /da/ to /ta/ along a VOT continuum. Stimuli fell into three distinct categories based on the pilot data described above. Near-Boundary (20 and 40 msec VOT) stimuli are poor examples of their phonetic categories, and because they are near the voiced-voiceless phonetic boundary should induce maximal competition between the voicing categories (-GoodFit, +Competition). Exemplar stimuli (-40 and 100 msec VOT) were good examples of their phonetic category and should be less subject to the influence of competition from the contrasting category (+GoodFit, -Competition)[2]. Extreme stimuli from the edges of the continuum (-100 and 160 msec VOT) were poor examples of their phonetic category, and

---

[2] It should be noted that in Experiment 1, the "best" exemplars in the set fell at 0 and 40 msec VOT, whereas in the current study, those best exemplars fell at -40 and 100 msec VOT. Because a full range of VOTs was not tested before choosing the stimulus set for Experiment 1, it may be the case that the 0 and 40 msec tokens simply represent the best exemplars in that stimulus range from 0 to 40 msec VOT. In the present study, -40 and 100 msec tokens were chosen as "exemplar" stimuli based on reaction time data from a pilot study which measured RTs for phonetic categorizations of a very wide continuum (-190 to +190 msec VOT). Some evidence exists that subjects may adjust their phonetic category space somewhat on the basis of the range of the with continuum which they are presented (Brady & Darwin, 1978; Keating, Mikos, & Ganong, 1981; Miller & Volaitis, 1989).

also are expected to show minimal effects of competition from the contrasting voicing category (-GoodFit, -Competition). It is predicted that anterior areas such as the left IFG should be driven by the influence of competition, and should thus show maximal activation for the Near-Boundary tokens, and in addition, should not show enhanced activation for the Extreme stimuli despite the fact that they are poorer exemplars of their phonetic category. It was also hypothesized that, given the responsiveness of the bilateral STG to the goodness of fit in Experiment 1, the STG should show greater activation for both Near-Boundary and Extreme stimuli compared to Exemplar stimuli.

3.3.2. Methods & Materials

3.3.2.1. Stimuli

Stimuli were those used in Pilot 3 and ranged from /da/ to /ta/. The VOT values of these stimuli were -100, -40, 20, 40, 100, and 160 msec. Based on previous pilot data, three of these (-100, -40, and 20 msec VOT) fell into the /da/ category, and the remainder fell into the /ta/ category. Stimuli from the ends of the continuum (-100 and 160 msec VOT) were referred to as ‘Extreme’ stimuli, Exemplar stimuli were -40 and 100 msec VOT, and Near-Boundary stimuli were those at 20 and 40 msec VOT.

In addition, two sine wave tones were generated at 720 Hz (“Low” tone) and 1250 Hz (“High” tone), which corresponded to the steady-state frequencies of F1 and F2 respectively for the syllable stimuli. These tones were equal in length (330 msec) to the syllable stimuli.

3.3.2.2. MR Participants

Fifteen subjects (7 males) participated in the MRI experiment for modest monetary compensation. Subjects gave informed consent in accordance with the Human Subjects policies of Brown University and Memorial Hospital of Rhode Island. Subjects ranged in age from 18 to 27 (mean age=23.2 ± 5.00). All subjects were strongly right-handed, as confirmed with an adapted version of the Oldfield Handedness Inventory (Oldfield, 1971). Subjects reported no known hearing loss or neurological impairment, and were screened for MR compatibility before entering the scanner.

3.3.2.3. Behavioral Procedure

The MR experiment consisted of four runs: two phonetic categorization runs and two tone categorization runs. During each of the phonetic categorization runs, subjects heard 20 repetitions of each VOT stimulus, for 40 total repetitions of each stimulus in the entire experiment. Likewise, during each tone categorization run, subjects heard 20 repetitions of each tone stimulus, for 40 total repetitions of each tone stimulus. Stimuli within each run were presented in a fixed, pseudo-randomized order. Stimulus presentation and response registration were carried out using the same apparatus described in Experiment 1. Subjects received alternating phonetic and tone categorization runs, starting with a phonetic categorization run.

Instructions for the phonetic categorization and tone tasks were the same as in Experiment 1. Subjects were familiarized with all stimuli during the anatomical scan, and were given 10 practice trials of each task type.

3.3.2.4. MR Imaging

Whole brain MRI was performed using the same apparatus and procedure as described in Experiment 1. As in Experiment 1, fifteen 5mm thick slices were acquired in each functional volume. Each functional volume acquisition lasted 1200 msec, followed by 20 msec of scanner-added time, plus 780 msec of added silence which was programmed into the acquisition sequence. As in Experiment 1, the result was a clustered design in which each TR consisted of 1200 msec of scanning, followed by 800 msec of silence, for a total TR of 2 seconds (see Experiment 1, Figure 12). Anatomical scanning took place as described in Experiment 1.

Auditory stimulus presentation was timed such that each stimulus fell in the center of the silent period between scans, starting 235 msec into the silent period. Stimuli were distributed into trial onset asynchrony (TOA) bins of 2, 4, 6, 8, 10 and 12 seconds using a Latin Squares design. The average TOA was 7 seconds. Each syllable run lasted 14 minutes, 8 seconds (422 EP volumes collected), and each tone run lasted 4 minutes and 48 seconds (142 EP volumes collected).

In addition, two functional volumes were automatically added to the beginning of each run to account for T1 saturation effects, but these volumes were not saved by the scanner. Stimulus presentation began four total volumes after the onset of each run to avoid saturation effects.

3.3.2.5. Data Analysis

3.3.2.5.1. Behavioral analysis

Behavioral data analysis was carried out as described in Experiment 1. Reaction time means and categorization data were collected and binned into the three syllable categories: Extreme, Exemplar, and Near-Boundary. Repeated-measures within-subjects

ANOVAs were carried out on the RT and categorization variables for syllables in these three categories.

3.3.2.5.2. MR analysis

3.3.2.5.2.1. Image Preprocessing

Image preprocessing was carried out as described in Experiment 1, with the exception of an additional slice timing correction step before motion correction. Slice time correction normalizes activation levels for slices acquired within a given volume by taking into account when during the volume acquisition that slice was acquired (Paradis, Van de Moortele, Le Bihan, & Poline, 2001; Van de Moortele et al., 1997). This was carried out by dividing each subject's concatenated functional dataset into individual runs. Next slice timing correction was run on each run separately, using the slice acquisition times given by the scanning protocol. Runs were re-concatenated, and motion-corrected as in Experiment 1. Functional datasets were resampled to 3mm isotropic voxels, transformed to Talairach and Tournoux space, and spatially smoothed as described in Experiment 1. These pre-processed datasets were then input to a deconvolution analysis.

3.3.2.5.2.2. Statistical Analysis

Deconvolution analyses were carried out as described in Experiment 1, with the exception that fourteen reference functions were entered into the analysis: six corresponded to the six VOT trial types, two for each tone type, and six for the motion correction parameters obtained in the motion correction analysis. Voxels which were not sampled for all fifteen subjects were masked from the final results. The raw fit coefficients for each subject and condition were then converted to percent change as described in Experiment 1.

A mixed-factor ANOVA was carried out on the raw fit coefficients with subjects as a random factor and stimulus condition as a fixed factor. The planned comparisons for this study were Exemplars vs Tones, Near-Boundary vs. Extreme, Near-Boundary vs. Exemplar, and Extreme vs. Exemplar. Statistical maps for each comparison were submitted to a clustering analysis which removed all activation which was not part of a cluster of at least 65 contiguous voxels activated at a voxel-wise level of $p<0.025$, which based on Monte Carlo simulations, yields a corrected cluster-wise threshold of $p<0.05$.

Regions of interest were defined based on areas which have previously been shown to be active in phonetic categorization tasks (see Experiment 1, Blumstein et al., 2005). These were the bilateral STG and Heschl's gyri, the left IFG, the left angular gyrus, the left insula, and the left and right cingulate gyri, and were defined using the AFNI interface to the Talairach and Tournoux atlas (Lancaster et al., 2000).

3.3.3. Results

3.3.3.1. Behavioral results

Categorization and reaction time data for all subjects run in the scanner is shown in Figure 15A, and the same data is shown binned by VOT category in Figure 15B. Subjects perceived stimuli categorically, showing near perfect consistency on all tokens except the 20 msec VOT token, which was identified as a /d/ 83% of the time. Reaction times for both the /d/ and /t/ ends of the continuum showed the expected pattern, with slower reaction times for stimuli near the boundary (20 and 40 msec VOT), and at the ends of the continuum (-100 and 160 msec) than those closer to the center of the phonetic category (-40 and 100 msec VOT). Results of an ANOVA performed on categorization data for stimuli binned into Extreme, Exemplar, and Near-Boundary categories revealed a main effect of VOT category on accuracy (F(2,28)=12.084, MSE=0.001, p<0.001). Post-

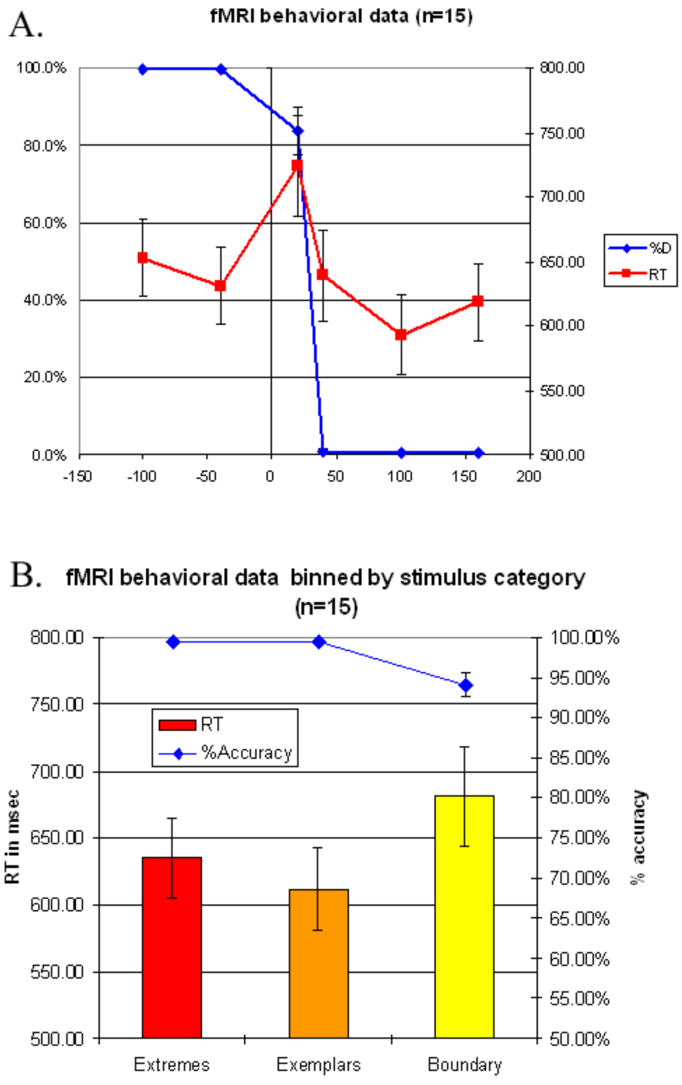

Figure 15. Behavioral data collected in the scanner (n=15). Error bars indicate standard error of the mean. Top (A) shows reaction time and % 'da' responses for all stimuli individually. Bottom (B) shows reaction time and % accuracy scores for three VOT bins.

hoc tests showed that this effect was due to a significant difference between Near-Boundary stimuli and both Extreme and Exemplar categories. No difference was observed between Extreme an Exemplar categories. An ANOVA performed on the binned RT means also revealed a significant effect of VOT category on RT (F(2,28)=15.535, MSE=1196.3, $p<0.001$). Post-hoc Newman Keuls tests revealed significant differences between the Near-Boundary stimuli and the other categories, but no difference between the Extreme and Exemplar categories.

3.3.3.2. Functional results

3.3.3.2.1. Cluster analysis

Clustering analysis was performed for all of the planned comparisons. Discussion will focus on comparisons between stimuli of different syllable types, namely Near-Boundary vs. Extreme, Near-Boundary vs. Exemplar, and Extreme vs. Exemplar. Clusters significant at a cluster-wise threshold of $p<0.05$ (65 contiguous voxels activated at $p<0.025$ or greater) are shown in Table 6.

Table 6. Areas of activation revealed in planned comparisons, thresholded at a voxel-level threshold of p<0.025, cluster-level threshold of p<0.05 (65 contiguous voxels). Coordinates indicate maximum intensity voxel for that cluster in Talairach and Tournoux (1988) space.

	Maximum Intensity			*Number of activated voxels*	*Local Maximum*
Anatomical Region	*x*	*y*	*z*		
Near-Boundary > Extreme					
L IFG, L MFG	50	-20	-1	897	0.1416
R IFG, R MFG	-47	-17	-4	621	0.1485
L/R Cingulate	2	-23	39	540	0.1314
R Superior Parietal Lobule, R IPL	-41	59	51	225	0.1332
Near-Boundary > Exemplar					
L IFG, L MFG, Postcentral (v. large)	53	-14	-4	949	0.2099
L/R Cingulate, Ant. Cingulate	2	-17	42	583	0.2018
R IFG, R Insula, R MFG	-47	-17	-4	518	0.2552
L/R Cuneus, post cingulate	-2	92	27	214	0.2856
R SPL, R IPL, R Cuneus	-26	68	54	201	0.1825
L SPL, L IPL	32	65	54	194	0.1087
L MTG, L STS, L STG	65	32	3	191	0.1215
R MTG, R STS, R STG	-62	38	3	114	0.0988
Extreme > Exemplar					
R MTG, STS, ant STG	-65	26	-1	119	0.183
L STG (ant), STS, Insula	62	11	9	78	0.1307
Post Cingulate	2	44	15	63	0.1406
Tones > Exemplar					
R IFG, R Insula, R Precentral	-47	-17	-4	351	0.1688
L Precentral, L Postcentral, L IPL	59	14	36	243	0.0934
L STG (ant), L Insula	59	2	9	125	0.0788
R Precentral, R IPL	-59	20	21	122	0.0946
L IPL	38	41	45	78	0.0752
L STG (anterior)	47	11	3	63	0.0852

3.3.3.2.1.1. Near Boundary vs. Extreme

All clusters active in this comparison were more active for the Near-Boundary than the Extreme condition. Activations were largely fell in frontal areas. There were two large clusters in the bilateral inferior frontal gyri, in both cases extending dorsally into the middle frontal gyri (Figure 16). Mean percent signal change plots of the three syllable conditions showed greater activation for Near-Boundary tokens than both Extreme and Exemplar stimuli, and little apparent difference between Extreme and Exemplar stimuli (Figure 16). Post-hoc Newman-Keuls tests revealed that in both areas,

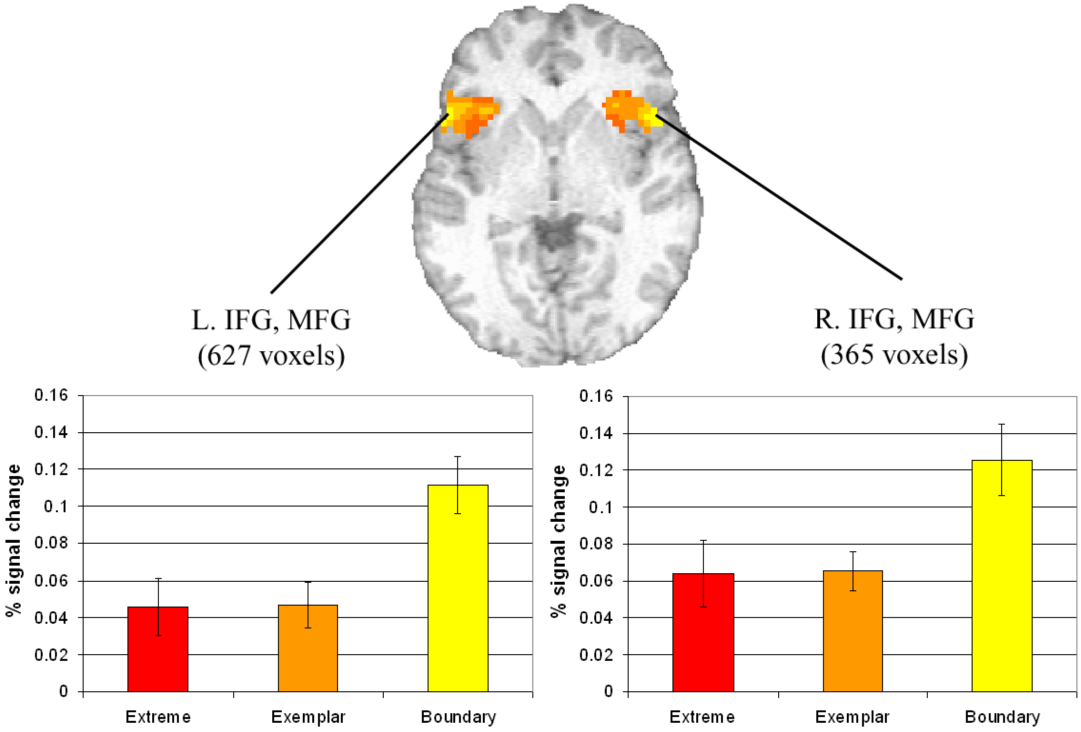

Figure 16. Left and right inferior frontal clusters active in the Near-Boundary vs. Extreme contrast. Axial slice shown at z=0, clusters active at a corrected threshold of p<0.05. Bar graphs show the extracted mean percent signal change for each cluster. Error bars indicate standard error.

the Near-Boundary stimuli had significantly greater activation than the other two categories (p<0.01), but there was no significant difference between Exemplar and Extreme categories.

In addition, there was a large midline cluster activated which was centered in the cingulate gyrus and extended into the anterior cingulate (Figure 17). As in the inferior frontal clusters, post hoc test revealed that this cluster showed significantly greater activation for Near-Boundary stimuli than the other two categories, and no difference between Extreme and Exemplar categories. Finally, there was a cluster in the right superior parietal lobule which extended into the IPL.

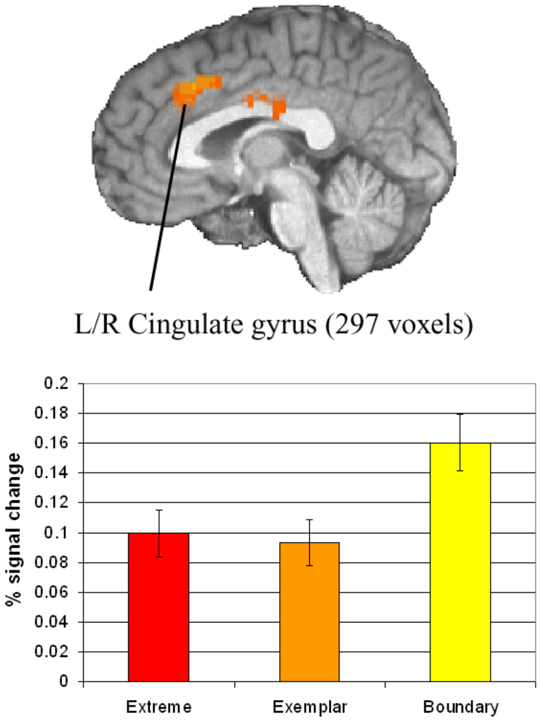

Figure 17. Cluster in the cingulate gyrus active for the Near-Boundary vs. Extreme comparison. Sagittal slice shown at x=0, clusters active at a corrected threshold of p<0.05. Bar graphs show the extracted mean percent signal change for each cluster. Error bars indicate standard error.

3.3.3.2.1.2. Near-Boundary vs. Exemplar

A number of clusters were activated for the comparison between Near-Boundary and Exemplar stimuli. All clusters showed more activation for Near-Boundary stimuli than Exemplar stimuli. Four of these clusters, in the left and right IFG, cingulate, and right superior parietal lobule, were nearly identical in placement and extent to those activated in the Near-Boundary vs. Extreme comparison.

Of interest, a cluster emerged which was centered in the left medial MTG, and which extended into the left superior temporal sulcus, and then into the left STG. A symmetric cluster was observed in the right MTG, STS, and STG as well (Figure 18). Mean percent change data extracted from both clusters revealed greater activation for both Extreme and Near-Boundary stimuli than Exemplar stimuli. Post hoc tests for mean percent change values in the left MTG cluster revealed significant differences between all stimuli. In the right MTG, the Exemplar stimuli were significantly less active than both the Near-Boundary and Extreme stimuli, but no difference emerged between the Near-Boundary and Extreme categories.

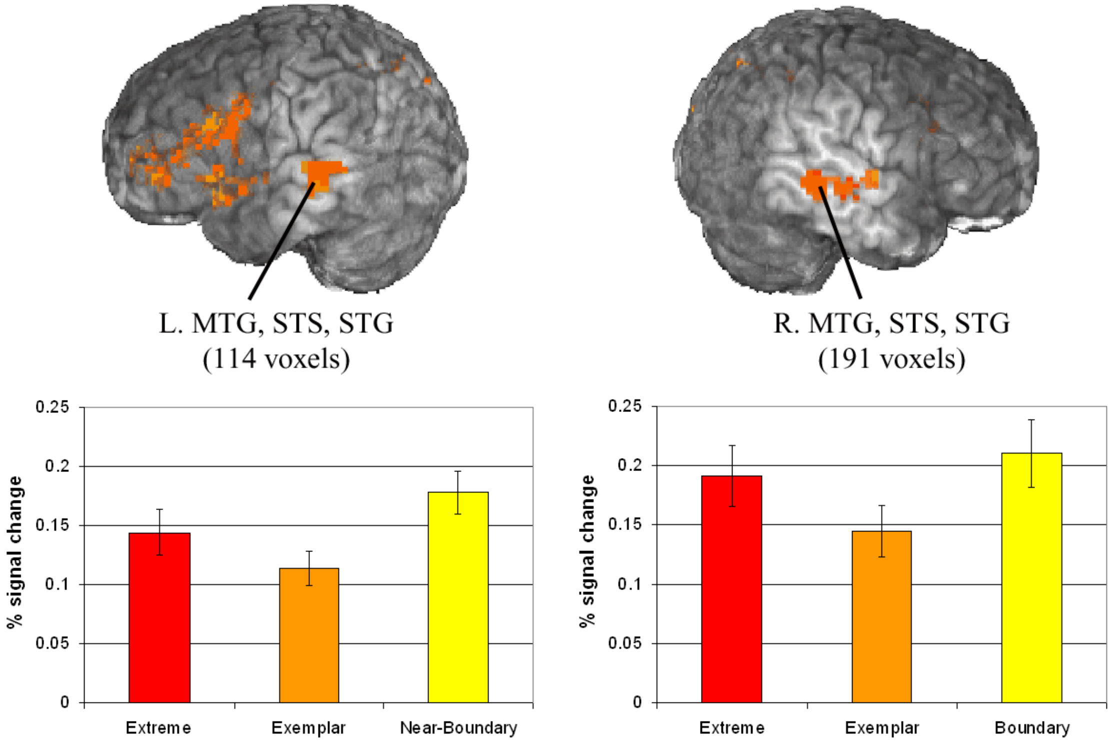

Figure 18. Left and right superior temporal clusters active in the Near-Boundary vs. Exemplar contrast. Clusters active at a corrected threshold of $p<0.05$. Bar graphs show the extracted mean percent signal change for each cluster. Error bars indicate standard error.

Significant activation also emerged in a midline cluster centered in the cuneus and extending into the posterior cingulate, and in the left superior parietal lobule, extending into the left IPL.

3.3.3.2.1.3. Extreme vs. Exemplar

As in the Near-Boundary vs. Exemplar comparison, two temporal clusters were activated in the comparison between Extreme and Exemplar stimuli, both with more activation for Extreme tokens. The right temporal cluster was very similar to the right MTG cluster defined in the Near-Boundary vs. Exemplar contrast. The left hemisphere cluster was centered in the left STG, and extended into the superior temporal sulcus and into the left insula, but did not reach the left MTG (Figure 19). Post hoc Newman-Keuls tests on the mean percent signal change revealed that the difference between Exemplar and Near-Boundary conditions was significant ($p<0.01$), while the difference between

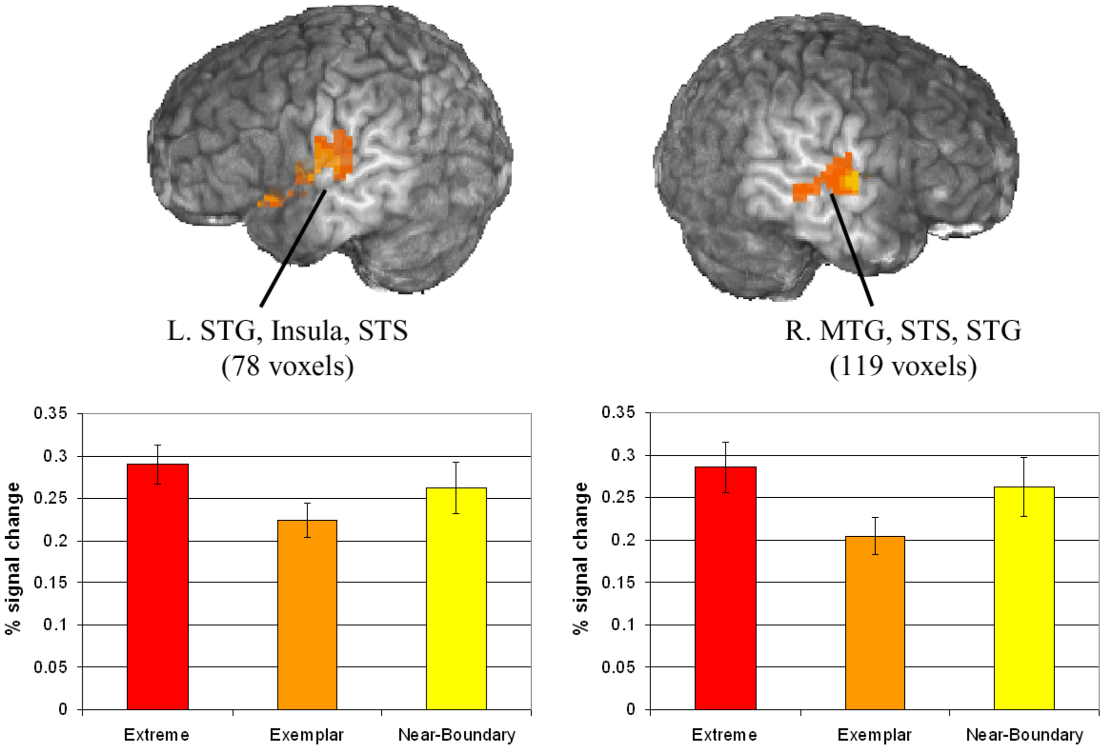

Figure 19. Left and right superior temporal clusters active in the Extreme vs. Exemplar contrast. Clusters active at a corrected threshold of p<0.05. Bar graphs show the extracted mean percent signal change for each cluster. Error bars indicate standard error.

Exemplar and Extreme tokens approached significance in the left temporal cluster. There was no significant difference between Extreme and Near-Boundary conditions. Likewise, post hoc tests in the right temporal cluster showed that Exemplar stimuli had significantly less activation than either Extreme or Near-Boundary stimuli, while there was no difference between Extreme and Near-Boundary stimuli.

3.3.3.2.2. ROI analysis

A region of interest analysis was performed to examine both posterior and anterior areas which had previously been implicated in phonetic categorization. This analysis largely corroborated the results found in the clustering analysis. As such, only

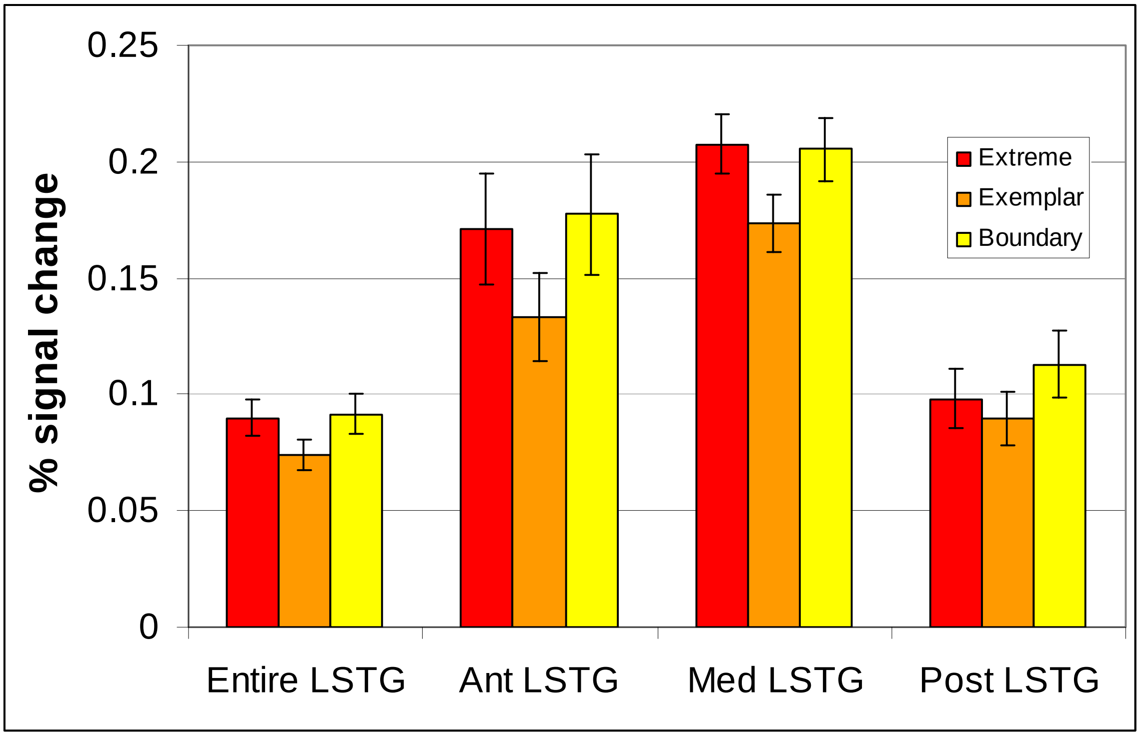

Figure 20. Region of interest analysis performed on the entire anatomically-defined left STG and sub-regions of the L STG. The anterior portion extended anteriorly from a plane at y=-8, the middle portion extended between planes at y=-8 and y=-35, and the posterior portion extended posteriorly from a plane at y=-36. Significant differences were observed in all areas except the posterior portion of the L STG. Error bars indicate standard error.

those regions which showed a different pattern of activation, or which did not emerge in the cluster analysis are reported in detail here.

Significant main effects of VOT category on activation levels were found in the left STG (F(2,28)=3.474, MSE=0.0004, p<0.045), in the anterior portion of the L STG (F(2,28)=3.554, MSE=0.002, p<0.042), in the medial portion of the L STG (F(2,28)=3.987 , MSE=0.001 p<0.030) (Figure 20). There was no significant difference between VOT categories in the posterior L STG, left or right Heschl's gyrus, the L MTG, left angular gyrus, or L insula. A significant difference between categories did emerge in the L IFG (F(2,28)=8.160, MSE=0.0004, p<0.002), but no main effect was found in the cingulate. The pattern of results in the L IFG replicated that found in the cluster analyses involving the Near-Boundary comparison.

Post-hoc Newman-Keuls tests were performed on the mean percent signal change values in the L STG as a whole, the anterior portion of the L STG, and the medial portion of the L STG. These tests revealed significantly less activation for the Exemplar condition than for the other two conditions in all areas, but no difference emerged between the Extreme and Near-Boundary conditions (Figure 20). Of interest, the pattern of activation was similar in the posterior LSTG, but there were no significant differences between stimuli.

3.4. Discussion

The goal of the present study was to decouple the relative effects of goodness of fit and competition on activation in anterior and posterior peri-sylvian areas as subjects performed a phonetic categorization task. Results were consistent with the hypothesis that inferior frontal areas are maximally sensitive to the competition between phonetic categories which arises especially when stimuli are near the phonetic category boundary. In contrast, superior temporal areas were shown to be sensitive to the goodness of fit of stimuli, regardless of whether they were close to the phonetic boundary (Near-Boundary) or far from that boundary (Extreme). These results are discussed in the context of current views of the functional neuroanatomy of speech processing.

3.4.1. Behavioral Results

The pattern of behavioral results for subjects run in the scanner was similar, but not identical to that found in Pilot 3. Reaction times increased as stimuli neared the phonetic category boundary, as was shown in Experiment 1. Such an increase may be a result of either the poor fit of these stimuli to the phonetic category or due to the competition which arises due to their proximity to the phonetic category boundary. Reaction times also increased for those stimuli which were poor examples of a phonetic category and were either extremely pre-voiced or had very long VOTs. This indicates that subjects were sensitive to the goodness of fit of a token to its phonetic category, even when the effects of competition were minimized. Unlike Pilot 3, however, reaction times were significantly longer for stimuli near the category boundary than those at the extreme ends of the continuum. This result is likely due in large part to the effect of the 10 msec VOT token, which showed much longer reaction times lower accuracy than any of the other tokens. Post-hoc inspection of the data from the voiceless end of the continuum alone showed that reaction times for Near-Boundary and Extreme stimuli were much more equivalent (40 msec VOT=638.68 msec ±35.0; 160 msec VOT=618 msec ±30.0) .

3.4.2. Functional Results

3.4.2.1. Competition and the inferior frontal gyri

As in Experiment 1, activation emerged in the left inferior frontal gyrus as a function of the degree of competition expected to arise from the contrasting phonetic category (Near-Boundary vs. Exemplar, Near-Boundary vs. Extreme). Unlike Experiment 1, however, this activation emerged in not only the left but also in the right inferior frontal gyrus. Importantly, increased activation did not emerge in frontal areas for poor tokens which were not near the phonetic category boundary when compared to good phonetic tokens (Extreme vs. Exemplar). These results suggest a role for the bilateral inferior frontal gyrus in resolving competition, and also indicate that this area is not generally sensitive to the goodness of fit of a token to its phonetic category. These results also provide evidence that activation in the inferior frontal gyrus is not strictly coupled to the general difficulty of the phonetic decision as measured by reaction time (Binder et al., 2004; Blumstein et al., 2005). Extreme stimuli were poor exemplars of their phonetic category, and as such showed increased reaction times when compared to Exemplar stimuli, but no concomitant increase in activation in the inferior frontal lobes was observed for the Extreme vs. Exemplar comparison.

Activation in the left inferior frontal gyrus often emerges when subjects are required to make an executive decision about a phonetic or phonological aspect of the incoming stimulus (see Poeppel, 1996a for review), and when subjects are required to make decisions about linguistic stimuli generally. Typically the pattern of activation is such that increased activation is shown for the more difficult condition or task. Thompson-Schill and colleagues (Kan & Thompson-Schill, 2004a; Thompson-Schill et al., 1997; Thompson-Schill, D'Esposito, & Kan, 1999) refocused this view and posited that the role of the left inferior frontal gyrus was in selecting between competing alternatives. This hypothesis was first proposed for semantic processing specifically, but recent evidence has shown involvement of an adjacent area, the left middle frontal gyrus, in coordinating the processing resources necessary for resolving the competition between lexical items with similar phonological structure (Prabhakaran et al., submitted) The pattern of activation observed in the present study supports the view that the left inferior gyrus is involved not simply in selecting between different semantic representations, but in resolving competition in linguistic tasks generally.

Of interest, competition-modulated activity emerged in this study in the right inferior frontal gyrus and surrounding areas, as well as in the more expected left-hemisphere region. Modulation of activation of the right inferior frontal gyrus is less reported in language tasks than is its left-hemisphere homologue, and its role in language processing is as yet unclear. Poldrack and colleagues (Poldrack et al., 2001) showed increasing activation in the bilateral IFG as auditory stimuli were compressed in time, presumably reflecting the increasing processing requirements necessary to understand the speeded speech. However, activation levels fell off when stimuli were so speeded as to be incomprehensible, which additionally suggests that the right IFG is sensitive to speech intelligibility, not simply to speech rate. A recent study in our lab investigating the discrimination of voice-onset time showed a compatible pattern of results (Hutchison, Blumstein, & Myers, submitted). Subjects engaged in a phonetic discrimination task to tokens which either fell within a phonetic category, or between two categories. Activation in the right inferior frontal gyrus was greater for discrimination of pairs which were acoustically similar than those which were more different, but only when those stimulus pairs could be reliably discriminated by subjects. Like the Poldrack study, these results suggest that the right IFG has a role in mediating the computational resources required for speech comprehension, but only when the task is one which may be consistently performed by subjects. Bilateral dorsolateral prefrontal cortex, of which the inferior and middle frontal gyri are subsections, has been also implicated in perceptual and conceptual selection (c.f. Kan & Thompson-Schill, 2004b and; Miller & Cohen, 2001 for reviews), in non-linguistic as well as linguistic tasks. Given that the phonetic categorization task used in this study may involve competition on both linguistic (phonetic) and non-linguistic (conceptual) levels, it is difficult with the present data to determine whether the activation observed in this study reflects the more domain-general selection role proposed for the bilateral frontal cortices, or whether it specifically reflects competition between linguistic alternatives.[3]

Results from this study clearly suggest that competition, and not goodness of fit in general, drives activation in the inferior frontal lobes. What remains unresolved is precisely what aspect of competition is responsible for increased reaction times to stimuli close to the phonetic boundary. In the phonetic categorization task, competition is inherent in several stages of processing. When subjects hear a token which falls near a phonetic boundary they must resolve the competition between the presented category and the competing phonetic category. This translates into increased competition in selecting between the two possible responses, /da/, and /ta/, and may in turn affect the difficulty of

---

[3] A post-hoc ROI analysis of data from Experiment 1 showed that the right IFG showed precisely the same pattern of activation as observed in the left IFG, with Boundary > Within-Category > Endpoint stimuli. However, these differences did not reach the clustering threshold for any of the comparisons between VOT stimuli.

the mapping between that chosen response and the motor response selected to carry out the button press. Thus, there is competition inherent in the phonetic decision-stage, and motor-planning aspects of the task for near-boundary tokens. If the inferior frontal lobes are involved specifically in executive aspects of language processing, activation for near-boundary stimuli might not be expected in tasks which require no overt response, for instance, in a passive listening task. However, if the inferior frontal gyri are involved in competition at the phonetic level of processing, competition-driven activation should be evident even when no overt response is required. Further research is necessary to test the hypothesis that the left inferior frontal gyrus is responsive to phonetic category competition in the absence of an overt response.

Evidence does exist suggesting that, the left inferior frontal gyrus is at least implicated in a variety of phonetic-phonological tasks which do require an overt response. The left inferior frontal lobe has been implicated in phonetic-phonological processing in a number of neuroimaging studies (c.f. Burton, 2001 for review; Joanisse & Gati, 2003; Poldrack et al., 2001; Poldrack et al., 1999). Perhaps more importantly, Broca's aphasics often show deficits in speech perception, with abnormal performance on both identification and discrimination tasks (c.f. Blumstein, 1995 for review; Blumstein, Baker et al., 1977; Blumstein, Cooper et al., 1977). Additionally, Broca's aphasics show elimination of semantic priming when the prime word ('pear') has an onset VOT which is decreased so it is closer to the phonetic boundary, and when that word also has a voicing competitior ('bear.') (Utman et al., 2001). Importantly, elimination of priming only emerges when phonetic-phonological competition (i.e. voiceless prime with a reduced VOT) interacts with lexical competition (i.e. 'pear' vs. 'bear'). Taken together, these results suggest a role for the left inferior frontal gyrus in resolving competition at multiple levels of linguistic processing.

3.4.2.2. Goodness of fit and the superior temporal lobes

Increased activation was seen in medial portions of the bilateral STG and MTG for stimuli which were poor exemplars of their phonetic category (Extreme, Near-Boundary) compared to those which were good exemplars (Exemplar). These results support the hypothesis that the superior temporal gyri are involved in the acoustic-phonetic processing necessary to map a stimulus to its phonetic category.

Previous results (Experiment 1) showed that, in a regression analysis, activation in the bilateral superior temporal lobes correlated with the goodness of fit of a stimulus to its phonetic category, but not with the overall difficulty of the phonetic decision as measured by reaction time. As mentioned previously, similar results were found by Binder and colleagues (Binder et al., 2004), who found that activation in the anterior superior temporal lobes was correlated with accuracy scores on a phonetic discrimination task. These accuracy scores were presumed to reflect the ‘perceptibility’ of stimuli presented in noise. In that study, accuracy and perceptibility were confounded, raising the possibility that the temporal activation observed was not the result of the increased processing required as the signal to noise ratio decreased, but to factors related to accuracy—for instance, response certainty or degree of attentional engagement. In the present study all stimuli with the exception of one token were perceived with near ceiling consistency. These results therefore support Binder’s claim that activation in the temporal lobes is related to perceptual processing rather than other factors. Curiously, the direction of correlation between activation levels and perceptual measures was opposite in the present study to the findings in the Binder et al. (2004) study. That is, Binder and colleagues showed increasing activation to tokens which were presented with a higher signal to noise ratio, and which would presumably require *less* processing to resolve their phonetic identity. In contrast, the results of the present study showed increasing activation to poorer tokens in a phonetic category, which should require *more* processing to resolve their phonetic identity. This difference in results cannot be resolved on neuroanatomical grounds—the clusters reported in the Binder study are anterior to those reported in this study, but there appears to be a large degree of overlap between the clusters reported by Binder and those identified in the current study.

One possibility for the difference between the two studies lies in the difference between the stimuli used. The Binder study used two tokens, ‘ba’ and ‘da’ which were both presumably good exemplars of their phonetic category. These tokens were presented in pairs in a variety of signal to noise ratios, thus the difficulty of the phoneme detection task used varied as a function of the perceptibility of the stimulus in noise, not as a function of difference processing resources needed for resolving variations in phonetic category structure. The superior temporal lobes have been shown to be sensitive to the acoustic properties characteristic of human speech when compared to other types of sounds (Belin et al., 2002; Belin et al., 2000; Zatorre et al., 2002). This being the case, an increase in activation may reflect the perceptibility the properties which characterize speech. In contrast, in the current study, all stimuli were similar with respect to the extent to which they typified human speech—that is, all were pronounceable tokens. As such, the activation pattern in the current study may reflect a baseline level of activity due to the temporal lobes’ responsiveness to human speech, which is then modulated by the goodness of fit of each token to its phonetic category. In sum, the superior temporal lobes seem to show sensitivity to tokens which deviate from the expected or prototypical stimulus, whether this deviation is the result of global changes in perceptibility due to presentation in noise (Binder et al., 2004) or due to the graded structure of phonetic categories (Blumstein et al., 2005; Guenther et al., 2004).

Another possible source of differences between the two studies may result from differences in the task. In the Binder study, subjects' task was always to decide whether a particular target syllable (either 'ba' or 'da', depending on the subject) was the first or second member of the pair. This task involves aspects of both phonetic identification and discrimination. When stimuli are presented in the clear, the decision can be made as soon as the first stimulus in the set has been identified, and hence is not dissimilar to the task in the current study. However, it is possible that as the signal to noise ratio decreases and stimuli become more difficult to identify, a subject might need to rely on comparison of the acoustic-phonetic properties of the two stimuli in the pair to determine which best matches the target token. As such, the increase in activation as stimuli become easier to identify may reflect a switch from discrimination to identification strategies. Unfortunately, no study to date has directly compared activation patterns for the same stimuli as a function of discrimination and identification tasks.

3.5. Conclusion

The current findings provide strong evidence for the role of anterior and posterior brain structures in mediating two aspects of the phonetic categorization process. Results from this experiment and from Experiment 1 together indicate that both anterior and posterior structures are sensitive to phonetic category structure. Both experiments support the conclusion that frontal structures such as the left and right IFG are involved in mediating competition between phonetic categories. The bilateral temporal lobes, in contrast, are sensitive to the goodness of fit of a stimulus to its phonetic category, and as such, have a role in resolving phonetic category membership, irrespective of the effects of competition. What is less clear is the extent to which these responses are fixed to the particular VOT values used in these studies, or whether they change based on the context in which a given phoneme is heard. Evidence from electrophysiology studies in human and non-human primates suggests that the reason that many of the world's languages divide voiced and voiceless stop categories at 20-30 msec VOT is that a particular group of auditory neurons can distinguish auditory events which are ~30 msec apart, but cannot resolve events which are closer together in time (Steinschneider et al., 1994, 1995). If this is the case, then the boundary which divides voiced and voiceless stops should be more or less fixed at 20-30 msec VOT. Nonetheless, evidence also exists that phonetic category boundaries vary in production as a function of speech rate (e.g. Miller, 1981b), and that perceptual boundaries may also vary as a function of lexical status, lexical frequency, and neighborhood density of the endpoints of the continuum (Connine, Titone, & Wang, 1993; Ganong, 1980; Newman, Sawusch, & Luce, 1997). It is an open question as to whether shifts in a phonetic category boundary in these situations reflect an inherent re-calibration of phonetic category space, or whether they are epiphenomena of the experimental tasks. The flexibility of phonetic space as a function of lexical status is the subject of the next study.

4. CHAPTER 4: EXPERIMENT 3, NEURAL BASES OF THE LEXICAL EFFECT

4.1. Introduction

Communication often takes place in a noisy environment. Acoustic masking of parts of words or variation in speakers' productions can result in local perceptual indeterminacies which require listeners to use available information to resolve the incoming acoustic stream. Evidence from phoneme restoration studies (e.g. Warren, 1970), results showing a lexical frequency bias for perception of speech in noise (Broadbent, 1967), and evidence of comprehension of grammatical vs. ungrammatical speech in noise (Miller, Heise, & Lichten, 1951) indicate that lexical, syntactic, and semantic information is used online as listeners to resolve indeterminacies when listening conditions are imperfect. What is less clear is at what stage or stages of processing this information comes into play. Does 'top-down' information influence basic perceptual mechanisms? Does it play a role during decision stages of processing? Or is top-down information used at both of these stages of processing?. Indeed, the top-down processing debate has been active in the literature for well over 20 years.

Of particular interest for this line of research is the effect of knowledge of the lexicon on the perception and identification of phonetic contrasts. One example of this sort of top-down effect is the 'lexical effect.' Ganong (1980) presented subjects with acoustic-phonetic continua varying in voice onset time (VOT), an acoustic cue to the perception of the phonetic feature of voicing in stop consonants. In half of the trials, the voiceless end of the continuum was a real word, and the voiced end of the continuum was a non-word (e.g. **peace**—beace). In the other half of the trials, the voiceless end of the continuum was a nonword and the voiced end of the continuum was a word (e.g. peef—**beef**). The presence of a meaningful lexical item at one end of the continuum was sufficient to shift the phonetic boundary of the continuum such that there were greater voiced responses to stimuli in the boundary range when the real word endpoint stimulus began with a voiced stop, and there were greater voiceless responses when the real word endpoint stimulus began with a voiceless stop. This lexically-biased shift in identification boundary or 'lexical effect' was taken by Ganong and others (see Pitt, 1995 for review) as evidence that lexical status influences perception of an acoustic-phonetic continuum in a direct way. The hypothesis that higher-level information (lexical status) directly influences the perceptual analysis of the acoustic properties of speech has not gone unchallenged, however. Others (e.g. Fox, 1984) have argued that the lexical effect is due to decision-related processes that result when subjects are forced to categorize ambiguous stimuli. This debate is, as of yet, unresolved in the literature (c.f. Connine and Clifton, 1987; Burton *et al.*, 1989; McQueen, 1991; Pitt & Samuel 1993). The inability to determine whether the lexical effect is perceptual or decision-related in nature may in part due to the limitations of the experimental methodology. The measures that experimenters have had at their disposal with regards to this question are those which have been associated with the phonetic categorization task itself: response, and response latency. Both of these measures are almost certainly influenced by both perceptual and decision-stage factors.

Neuroimaging techniques, particularly event-related fMRI, offer a new opportunity to address the question of the basis of lexical effects on phonetic categorization. Specifically, it may be the case that activation patterns due to lexical effects emerge in those brain areas which have been previously implicated in lower-level speech processing. Such a pattern may be taken as evidence that lexical status influences perception in a direct way. Alternatively, it may be the case that lexically modulated activation patterns only emerge in frontal areas which have been associated with executive processes. Such a case would lend credence to the argument that lexical effects operate uniquely on the level of decision-stage processes. Previous work (Blumstein et al., 2005) has shown sensitivity to phonetic category structure in perceptual areas such as the superior temporal gyrus (STG) as well as in areas linked to executive components of language comprehension, such as the left IFG, and those linked to decision-making in general, such as the anterior cingulate. In this study, this sensitivity is exploited to determine whether lexically-biased shifts in phonetic category boundary are perceptual or decision-related in nature.

Both neuroimaging data and lesion studies commonly implicate two neural areas involved in speech processing, the bilateral superior temporal gyri (STG) and the left inferior frontal gyrus (IFG) plus adjacent areas, which includes Broca’s area. Classical lesion models of language processing (Geschwind, 1965) and neuroimaging data (c.f. Binder & Price, 2001; Hickok & Poeppel, 2000; Scott & Johnsrude, 2003) support a neural circuit which proceeds from projections from the medial geniculate nucleus to Heschel’s gyrus (primary auditory cortex), to adjacent superior temporal areas, then anteriorly to inferior frontal structures such as the operculum, IFG and insula. The proximity of the STG to primary auditory cortex makes it a prime candidate for early acoustic and phonetic processing. Indeed, bilateral STG are activated when subjects are engaged in phoneme discrimination tasks (Zatorre et al., 1992) and phoneme detection tasks (Demonet et al., 1992; Demonet, Price, Wise, & Frackowiak, 1994). However, activation in these areas does not appear to be significantly modulated by task difficulty (Binder et al., 2004; Blumstein et al., 2005), and is not correlated with reaction time performance. However, these areas are sensitive to the acoustic and acoustic-phonetic properties of the stimulus (Binder et al., 2004; Blumstein et al., 2005; Giraud & Price, 2001, c.f. also Experiment 2) and presentation rate (although see Fiez et al., 1995; Mummery, Ashburner, Scott, & Wise, 1999; Price et al., 1992). The sensitivity of the STG to speech-relevant properties of the stimulus is highlighted by a number of studies have shown that STG activation increases as auditory stimuli become more speech-like (Belin et al., 2002; Belin et al., 2000; Binder et al., 2000; Giraud & Price, 2001; Zatorre & Belin, 2001). The relative insensitivity of the STG to task demands and its sensitivity to speech sounds support a role for the STG in lower-level perceptual and acoustic processing.

In contrast, activation in inferior frontal areas appears to be significantly affected by task demands. While activation in STG is reported for almost every study involving auditory speech stimuli, activation in inferior frontal areas has been less consistently reported (Poeppel, 1996a, 1996b). When it is reported, frontal activation tends to be left-lateralized (but see Paulesu et al., 1993; Poldrack et al., 1999). IFG activation occurs when subjects are required either to hold stimuli in memory for a period of time, when they are required to monitor the stimulus stream for a particular phoneme (e.g. Demonet et al., 1992; Demonet et al., 1994), or when discrimination tasks involve segmentation of the phoneme from its context (Burton et al., 2000; Zatorre et al., 1992) Finally, studies have shown that activation in the L IFG correlates with the difficulty of the phonetic decision, and in many cases is tightly correlated to reaction time measures (Binder et al., 2004; Blumstein et al., 2005). The activation of the left IFG in meta-linguistic tasks such as phoneme identification and phoneme monitoring as well as under conditions of competition and stimulus uncertainty supports a role for these areas in phonetic decision-stage processes.

A final area which appears to be recruited generally in decision-stage processing is the anterior cingulate gyrus (AC) (Gehring & Knight, 2000). This area is activated when subjects are engaged in tasks which involve competition between responses (Carter et al., 1998) and for experimental trials on which subjects make errors (e.g. Rissman et al., 2003), independent of the stimulus modality. This area has also shown modulation of activation as a function of phonetic category structure, with increasing activation as stimuli approach the phonetic boundary (Blumstein et al., 2005). Thus, the anterior cingulate appears to be involved in decision-stage processing in language tasks as well as other experimental tasks.

If the lexical effect reflects early use of lexical information in guiding speech perception, it is expected that the lexical effect will influence basic perceptual mechanisms (STG) as well as decision mechanisms (IFG and AC). As in Experiments 1 and 2, it is expected that modulation of activation in the STG, left IFG and AC will emerge as a function of the proximity of the stimulus to the phonetic category boundary. Given that the phoneme categorization task used in the present study is an explicit task, certain executive demands are necessarily imposed. As in Experiments 1 and 2, these are expected to vary as a function of the ambiguity of the stimulus. As such, more activation should emerge in the L IFG and AC within the same VOT value for stimuli which fall on the category boundary, and are therefore more ambiguous, than those which do not.

Modulation of activation within executive regions as a function of the lexical effect is expected to emerge whether the lexical effect itself is perceptual or decision-related in nature, due to the inescapable fact that the task itself involves an executive decision. However, as described above, modulation of activation should occur as a function of the indeterminacy of the stimulus and hence the difficulty of the executive decision. However, similar changes in activation in STG due to the biasing presence of a lexical item may be taken as evidence that lexical effects operate at the perceptual level. As described above, the STG has been shown to be responsive to the characteristics of the speech stimulus, with increasing activation as a function of the ‘goodness of fit’ of the stimulus to a phonetic category (Experiments 1 & 2). It is also relatively immune to factors such as task difficulty which are known to affect frontal and midline structures (Binder et al., 2004; Blumstein et al., 2005). If the influence of the lexical effect is to fundamentally alter the phonetic category boundary, sensitivity to the ‘new’ boundary should be seen within the STG. Specifically, more activation should be seem for the same VOT value when that value falls on the phonetic category boundary than when it does not.

Some evidence exists to support the notion that the neural substrate of the lexical effect is posterior, and potentially temporal, rather than anterior. A study with Broca’s and Wernicke’s aphasics demonstrated that Broca’s aphasics, who have damage in the left IFG and other anterior areas, show an abnormally large lexical effect, while Wernicke’s aphasics, who have damage to left STG and other posterior structures, show no lexical effect (Blumstein, Burton, Baum, Waldstein, & Katz, 1994). One possible interpretation of this data is that lexical effects arise due to interaction between the STG and adjacent areas including the middle temporal gyrus and supramarginal gyrus which are known to be involved in lexical processing (Binder & Price, 2001). When these areas are damaged, as they typically are in Wernicke's aphasia, the lexicon may not influence perception. In contrast, when these areas are intact, but inferior frontal areas are damaged, as in Broca's aphasia, lexical effects strongly influence perception.

The aim of the present study is to investigate the neural basis of the lexical effect, and to use these results to inform the debate over whether such effects are perceptual or decision related. To this end, subjects were presented with lexical effect continua which varied in the phonetic feature of voicing. That is, the continua varied in VOT with the endpoint stimuli of the continua corresponding to a word at one end of the continuum and a nonword at the other end of the continuum. Activation was measured for each stimulus type, and statistical comparisons were made between continua, but within the same VOT value. These comparisons were made to determine the neural area(s) where modulation of activation occurs as a function of lexically biased shifts in phonetic boundary. Such modulation of activation is expected in the left IFG and cingulate; modulation of activation in the STG would provide evidence that lexical effects have a perceptual component as well.

4.2.General Methods: Stimulus Generation

Candidate word/non-word pairs were identified which differed only with respect to the voicing of the initial stop consonant (e.g. gift-*kift*). All pairs were one-syllable, and conformed with a [Stop V C (C)] pattern. These words were identified via a search through the MRC psycholinguistic database (Coltheart, 1981). Pairs that began with alveolar stops (/t/ or /d/) were avoided, as this place of articulation has been shown to be problematic in studies of the lexical effect (Burton & Blumstein, 1995; Newman et al., 1997; Pitt & Samuel, 1993). Each pair in which the voiced member of the pair was a real word was in turn linked to a word/non-word pair in which the voiceless member of the pair was a real word to create a potential stimulus set. Word/non-word sets shared the same vowel, and if possible, the same manner of articulation in final consonant(s), (e.g. gift-*kift* with *giss*-kiss). These linkages were made so that within each set of word/non-word pairs, word endpoints had similar written frequency (Kucera & Francis, 1967), and word and non-word members of each pair were matched for neighborhood density (Luce & Pisoni, 1998), and phonotactic probability (Vitevitch, Luce, Pisoni, & Auer, 1999) insofar as possible. After this process, one potential stimulus set survived the selection criteria: and gift-*kift* with *giss*-kiss (see Table 7 for a list of the lexical properties of the final stimulus set).

The stimulus set was digitally recorded in list intonation by a male native speaker of American English. These recordings were digitally transferred to a Dell computer and resampled to 22.050 kHZ. Stimuli were trimmed so that there was no silence at the beginning or end of each sound file.

Four eight-point voicing continua were created for each set using the BLISS speech editing program (Mertus, 1989). Each continuum was base-generated from the voiced, /g/ end of the continuum, and modified by splicing in successive amounts of aspiration from the /k/ sound taken from the voiceless token. For example, the ‘gift’ end of the gift-*kift* continuum was created by splicing together the initial /k/ burst from ‘kift’, plus the first quasi-periodic pitch period before the vowel from ‘kift’, plus the entire /Ift/ portion from ‘gift.’ Successive tokens along the continuum were created by removing one pitch period at a time from the beginning of the vowel, and replacing it with an equal amount of aspiration noise which followed the burst of the /k/ in ‘kift.’ In this example, this process yielded 8 stimuli with VOTs of 3, 11, 18, 25, 33, 40, 47, and 54 msec, all of equal length (528 msec). In order to create the other continuum in the set, *giss*-kiss, the initial stop plus vowel portion (135 msec) of each stimulus on the gift-*kift* continuum was spliced to 393 msec of the /s/ offset from ‘giss,’ yielding stimuli with a length of 528 msec. Thus, the two continua shared the same onsets, and hence VOT values, and all tokens in both continua were the same length. This process was repeated, but this time the base used to generate all tokens was taken from the ‘giss’ token rather than the ‘gift’ token, resulting in four VOT continua: gift-*kift* and *giss*-kiss generated from ‘gift’, and gift-*kift* and *giss*-kiss generated from ‘giss.’ Stimuli were checked for naturalness and waveforms were visually examined for acoustic discontinuities.

4.3. Pilot Experiment: Confirming the presence of a lexical effect

Preliminary pilot work showed that the category boundary for both the giss-kiss and gift-kift continua fell roughly between stimulus 5 and 6, yielding many more perceived /g/ sounds than /k/ sounds. In order to approximately equate the proportion of /g/ and /k/ responses, the first stimulus in the continuum (3 msec VOT) was removed, leaving a seven-point VOT continuum ranging from /g/ to /k/.

Table 7. Properties of the stimulus set used in the pilot and fMRI experiments. "Frequency" indicates written frequency (Kucera & Francis, 1967), FWND indicates frequency-weighted neighborhood density, (Luce & Pisoni, 1998), and mean phoneme and biphone probabilities are calculated after Vitevitch & Luce (2004).

Stimulus /phonology/	*Frequency*	*FWND*	*Mean Phoneme Probability*	*Mean Biphone Probability*
gift /gIft/	**33**	**9.531**	**0.058**	**0.002**
kift /kIft/	--	11.794	0.075	0.002
kiss /kIs/	**17**	**30.359**	**0.089**	**0.009**
giss /gIs/	--	26.366	0.067	0.009

A pilot experiment was conducted in order to ensure that subjects categorically perceived items along the voicing continuum, and that a majority of subjects showed lexically biased shifts in phonetic category boundary. In addition, the basis of the generated continua (‘gift’ or ‘giss’-based) was examined, to ensure that any differences found between continua were consistent regardless of the basis of the set. Stimuli consisted of 20 tokens of each of the seven points on each continuum. There were four continua, gift-*kift* and *giss*-kiss generated from ‘gift’, and gift-*kift* and *giss*-kiss generated from ‘giss.’ Recall that continua generated from the same base stimulus (‘giss’ or ‘gift’) shared the same VOT values as well as the same overall length. As such, the crucial comparison was always between continua which were generated from the same base stimulus. Stimuli were presented in a randomized order, and were blocked by continuum.

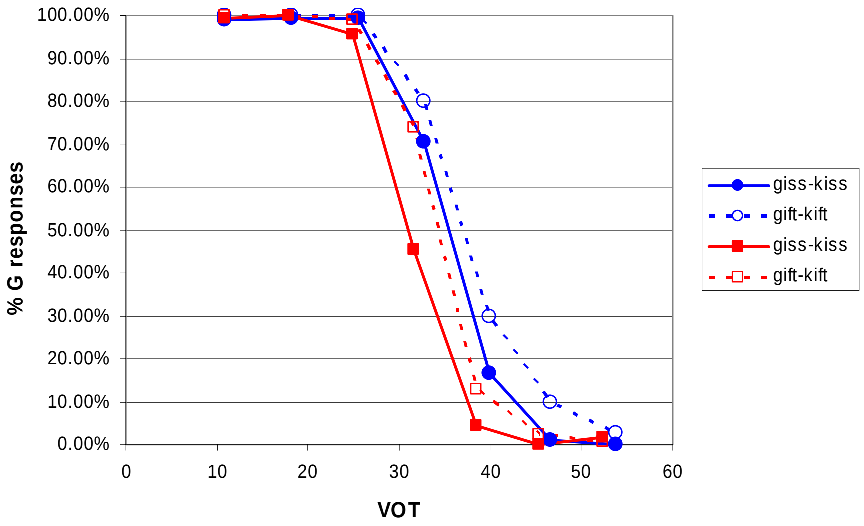

Figure 21. Categorization data from nine pilot subjects. Blue lines (circles) indicate 'gift'-based continua; red lines (squares) indicate 'giss'-based continua. VOT displayed in msec.

Nine pilot subjects were asked to listen to each stimulus, and to identify whether each started with a /g/ sound or a /k/ sound, and to press a corresponding button as quickly and as accurately as possible with their dominant hand. Reaction time and accuracy data were collected, with reaction times measured from the onset of the stimulus.

In general, subjects perceived stimuli categorically, with highly consistent identification of most tokens along the continuum (Figure 21). In addition, small lexically-biased shifts in phonetic category boundary were shown in both the 'gift'-based and 'giss'-based continua. Identification functions were plotted for each subject and for each continuum. Boundary VOT values were calculated for each continuum. First, percent /g/ scores for each point on the continuum were converted to z-scores. A critical range of VOT values was then selected which encompassed the minimum and maximum z-score values (for nearly all subjects, 5 to -5, or 100% consistent identification). In cases where a z-score value before the maximum or after the minimum deviated from that maximum or minimum, the score was only included if it deviated by more than 5% from that maximum or minimum. The selected z-scores for each continuum were then submitted to a linear regression, which output slope and y-intercept values. The x-intercept (the putative phonetic category boundary value) could then be calculated using the formula: x intercept = -(y intercept)/slope.

Seven of nine subjects showed a boundary shift with longer VOT boundary values for '*giss*-kiss' than 'gift-*kift*', within the 'gift'-based set. Eight of nine subjects showed a boundary shift in the expected direction within the 'giss'-based set (Table 8). Boundary values for each subject were submitted to a 2-way repeated measures ANOVA with Continuum ('*giss*-kiss' or 'gift-*kift*') and Basis (based on 'gift' or based on 'giss) as factors. There was a significant main effect of Basis (F(1,8)=22.687, p<0.001), with the 'gift'-based continuum overall having longer VOT boundaries than the 'giss'-based continuum. There was also a significant main effect of Continuum (F(1,8)=6.133, p<0.038), with the 'gift-*kift*' continua having longer VOT boundary values, as expected. There was no significant interaction (F(1,8)=0.6).

Table 8. Phonetic category boundary in msec of VOT. Boundary shift size was calculated by subtracting the 'giss-kiss ' boundary value from the 'gift-kift' boundary value.

	Based on 'gift'			Based on 'giss'		
Subject	*giss*-kiss	gift-*kift*	Shift Size	*giss*-kiss	gift-*kift*	Shift size
GK_01	38.48	36.55	-1.93	28.71	36.84	**8.13**
GK_02	36.35	45.59	**9.24**	28.97	38.01	**9.04**
GK_03	42.78	42.26	-0.52	40.96	29.11	-11.85
GK_04	30.04	32.55	**2.50**	31.71	35.07	**3.36**
GK_05	35.08	41.93	**6.86**	31.49	38.09	**6.61**
GK_06	39.20	42.33	**3.13**	31.90	41.22	**9.32**
GK_07	32.61	38.48	**5.87**	27.91	34.33	**6.42**
GK_08	35.85	36.03	**0.19**	28.55	31.34	**2.79**
GK_09	36.11	36.55	**0.44**	28.86	34.65	**5.80**
Mean	36.28	39.14	**2.86**	31.01	35.41	**4.40**

Although the magnitude of the lexical effect did not significantly vary between the two base conditions, the continua generated from the 'giss' token showed on average the largest lexical effect, and hence these two continua were selected for use in the fMRI experiment. Stimuli had VOTs of 11, 18, 25, 32, 38, 45, and 53 msec, and all stimuli were 528 msec in length. These stimuli will be referred to by the stimulus number corresponding to their place along the particular VOT continuum with stimulus 1 representing the first stimulus at the voiced end of the continuum and stimulus 7 representing the last stimuli at the voiceless end of the continuum (i.e. gift1—gift7, and giss1—giss7).

4.4. Experiment 3A: Non-word to Non-word Continuum

4.4.1. Introduction

While the pilot experiment established that the selected stimulus set and procedure were likely to result in lexically-biased shifts in phonetic category boundary, a question still remained as to whether this shift was made manifest by a shift of both continua or only one continuum. In theory, both continua could show shifts in the phonetic category boundary due to the biasing presence of a lexical item, with the giss-kiss boundary moving closer to the voiced end of the continuum, and the gift-kift boundary shifting towards the voiceless end of the continuum. In both cases, the shift would yield more phonetic categorizations corresponding to real words.

Previous studies of the lexical effect have not directly examined whether each of the two continua used to demonstrate the lexical effect show changes in the locus of the phonetic category boundary or whether the lexical effect is attributable to only one continuum showing a significant change in the locus of the boundary. To explore this question, it is necessary to compare the two lexical effect continua to either a baseline nonword-nonword continuum or a baseline word-word continuum. Fox (1984) asked subjects to perform a phonetic identification task on a word—word continuum (e.g. baddad) and a non-word—non-word continuum (e.g. /bae/ to /dae/) as well as on two lexical effect continua (e.g. bag-dag, bab-dab). Statistical comparisons were made between the lexical effect continua boundaries, and a separate comparison was made between the word—word and non-word—non-word continua. However, no direct statistical comparison was made between the boundaries for the lexical effect continua and those for either the word—word or non-word—non-word continua. In study by Miller & Dexter (1988) investigating the effect of perceived speech rate on the lexical effect, a word—word (bee—pea) continuum was presented to subjects in addition to the lexical effect continua (beace-peace; beef-peef). The boundary for the word-word continuum was compared to the average boundary value across both lexical effect continua in three RT bins (fast, medium, and slow), but was not compared to either continuum individually in any of the bins. Hence, it is difficult to determine from existing data whether both lexical effect continua move with respect to a lexically unbiased continuum, or whether only one does.

This question impacts the present study in a significant way: if only one continuum shows a significant shift in phonetic category boundary, only that continuum may be said to be subject to biasing lexical factors. In such a case, one might expect activation differences between the two continua to be asymmetric as a reflection of this fact. However, if both continua show significant shifts in the locus of the phonetic category boundary, then activation differences between the two continua should be approximately symmetric. That is, if there is a significant difference in a given area for the comparison [VOT value which falls on the gift—kift boundary] – [same VOT value in the giss—kiss continuum], there should also be a significant difference in the same area for the comparison [VOT value which falls on the giss—kiss boundary] – [same VOT value in the gift—kift continuum]. In order to investigate this question, a study was run in which a non-word to non-word continuum was used. The non-word to non-word continuum should show no effects of lexical bias on the phonetic category boundary and hence serves as a neutral baseline against which the lexical effect continua can be compared. Thus, the current experiment will investigate whether the lexical effect emerges in both lexical effect continua by comparing the locus of the phonetic boundary in each of the lexical effect continua obtained in the pilot to a baseline phonetic boundary obtained from a non-word—non-word continuum.

4.4.2. Materials and Methods

A nonword to nonword continuum, (gish-kish) was created by taking 136 msec of the onset portion (stop plus vowel) of the ‘giss’- based stimuli generated for the pilot and fMRI experiment, and splicing onto that onset 392 msec of the sound /sh/, taken from a recording of the same male speaker saying the word /gIsh/. In this way, a non-word to non-word (gish—kish) continuum was created which had the same onsets and length as the set used in the pilot experiment. These data were compared to the ‘giss’-based continua from the pilot experiment.

4.4.3. Results

Ten subjects participated in Experiment 3A. The experimental design and task were the same as for the pilot experiment. Identification functions for Experiment 3A are displayed together with those from the ‘giss’-based continua from the pilot in Figure 22. The mean phonetic category boundary for the nonword-nonword continuum was 29.63 msec (SD=2.19). A one-way ANOVAs on individual subject category boundaries

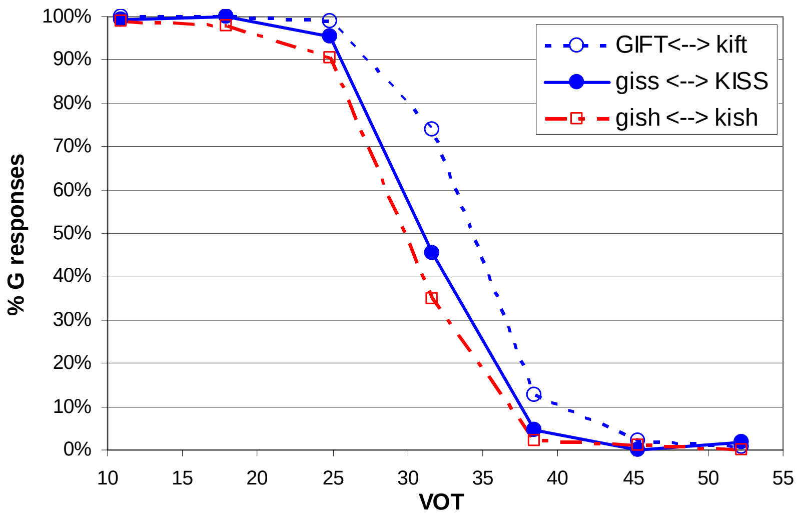

Figure 22. Word—Non-word 'giss'-based continua from Pilot 1 (blue lines, n=9) contrasted with Non-word—Non-word continuum from Experiment 3A (red line, n=10).

indicated that the boundary for gish—kish was not significantly different from the giss—kiss boundary (F(1,18)=0.879, MSE=8.965, p<0.362), but was significantly different from the gift—kift boundary (F(1,18)=17.808, MSE=158.076, p<0.001). This data suggests that of the two lexical effect continua, only the gift—kift continuum shows a significant shift away from a non-biased baseline continuum.

4.4.4. Discussion

A number of potential explanations exist for the failure of both lexical effect continua to show a significant shift in the locus of the phonetic boundary. It is important first to note that the failure to show a boundary shift in the giss-kiss continuum is unlikely to be due to factors having to do with the stimulus set itself. 'Kiss' is a noun of fairly high lexical frequency (17 occurrences per million, see Table 7), and thus should be easily recognized as a word during the identification task. Studies using stimuli with lower lexical frequency such as duke (freq=11)—tuke and doot—toot (freq=3) have found significant lexical effects (Burton, Baum, & Blumstein, 1989). Previous work (Newman et al., 1997) has shown using only non-word—non-word continua that changes in the phonetic category boundary can be induced as a function of the lexical density of the endpoints of the continuum, with the boundary shifting away from the higher-dense end of the continuum, hence resulting in a bias towards high density neighborhoods. Given that ‘kiss’ has a higher density than does its non-word counterpart, ‘giss’, the boundary shift for this pair should be towards the non-word end of the continuum, which is the same direction a lexically-biased shift would move in. Hence, lexical density cannot explain the lack of a lexical shift.

One possibility for the lack of shift for the giss—kiss continuum is that the acoustic space that distinguishes voiced and voiceless stops is asymmetric. That is, while most subjects have a voicing boundary that falls at about 20-30 msec VOT, the range of VOTs that characterizes the majority voiced tokens is much smaller (about -40 to 20 msec VOT) than is the range that characterizes voiceless tokens (about 30 msec to 170 msec VOT) (Lisker & Abramson, 1964). Hence, there is a greater range of values at the voiceless end of the continuum than at the voiced end, which would result in greater ease in moving a phonetic boundary towards the voiceless end of a continuum than towards the voiced end. Some evidence exists supporting the notion that there may be a lower bound on the movement of the VOT boundary towards the voiced end of the continuum. Evidence from production has shown that, when speech slows, VOTs for voiceless stops become considerably longer, while they remain the same for voiced stops (Miller, Green, & Reeves, 1986), which suggests that there is less room for movement within the voiced category than within the voiceless category.

Miller & Dexter (1988) showed that, when subjects' lexical effect categorization data is divided into thirds based on three reaction time bins (fast, medium, and slow), the size of the lexical effect diminishes as subjects respond more quickly. However, this diminution of the effect is manifested by all continua converging on the short (voiced) VOT boundary values. The authors' interpretation of this data is that when subjects respond very quickly to stimuli, they ignore later-occurring information in the word, and in essence treat the word as if it were shorter, and hence spoken at a faster rate than it actually was. Speech rate is known to be an important factor influencing phonetic category boundaries for VOT continua (Miller, 1981a), with faster perceived speech rates producing shorter VOT boundary values. An alternative explanation of these data based on our results is that, while speakers may treat words as if they were spoken at a faster rate when they respond quickly, and the overall shortening of the VOT boundary may be a result of this, the reason the lexical effect continua converge is because one continuum is essentially already at the 'floor' for the voicing boundary range.

Whatever the ultimate explanation for this phenomenon, it is important to note that, while both continua have the *potential* to be influenced by lexical factors, only one continuum actually shows behavioral evidence of a lexically-biased boundary shift. This raises an important question for interpretation of fMRI data with respect to these two continua. If, for instance, the important driving factor behind brain activation is the potential for lexical factors to influence the phonetic category boundary, the two continua should be very similar with respect to their pattern of activation. However, if activation patterns are driven by the 'usefulness' of lexical information in influencing phonetic category boundary, different patterns of activation should emerge for the gift—kift stimuli than for the giss—kiss stimuli.

4.5. Experiment 3B: Neural Bases of the Lexical Effect

4.5.1. Introduction

This study is designed to investigate whether lexically-modulated activation emerges in perceptual or decision-related neural areas. Subjects performed a phonetic categorization task on stimuli from the two seven-point continua from the pilot experiment, gift—kift and giss—kiss. Patterns of activation were compared across both continua for the same VOT value, with the crucial comparison being where the VOT value for one continuum falls on the phonetic category boundary and the analogous VOT value for the other continuum is no longer on the phonetic category boundary, having been affected by a lexically-induced boundary shift. Modulation of activation is expected based on sensitivity to this shift. In particular, increased activation should emerge for the boundary stimulus in executive regions such as the anterior cingulate and the left IFG due to the increased executive demands of processing ambiguous stimuli, with more activation for stimuli which fall on the phonetic boundary than those that do not. In addition, increased activation may be seen in the bilateral STG in response to boundary stimuli due to the demonstrated sensitivity of the STG to the ‘goodness of fit’ of a token to its phonetic category.

4.5.2. Materials and Methods

4.5.2.1. Stimuli

The seven-point gift—kift and giss—kiss continua used in the pilot experiment were used for the fMRI experiment (for lexical properties of the stimuli, see Table 7, and for stimulus parameters see Table 9). All stimuli were 528 msec in length. A tone control task was also developed. To this end, two sinewave tones corresponding to the 1st and 2nd formant frequencies of the steady-state portion of the vowel (457 and 1524 Hz) were created with durations of 528 msec.

Table 9. Properties of the stimuli used in the fMRI experiment

Stimulus	*VOT*	*Vowel length*	*Offset length (/ft/ or /s/)*
Gift1/Giss1	11	124	393
Gift2/Giss2	18	117	393
Gift3/Giss3	25	110	393
Gift4/Giss4	32	103	393
Gift5/Giss5	38	97	393
Gift6/Giss6	45	90	393
Gift7/Giss7	52	83	393

4.5.2.2. MR Participants

Nineteen participants (eleven females) took part in the MR version of the experiment for modest monetary compensation. Participants ranged from 19 to 51 years of age (mean age=27 yrs), and were all right handed, as indicated by a modified version of the Oldfield handedness inventory (Oldfield, 1971). Participants all reported having normal hearing, and had no known history of neurological disease. All participants gave written informed consent according to guidelines established and approved by the Human Subjects Committees of Brown University and Memorial Hospital of Rhode Island, and were screened for MR safety before entering the scanner room. Two of the nineteen subjects were eliminated from the study because they failed to show the lexical effect (see Behavioral Results).

4.5.2.3. Behavioral Procedure

The MR experiment consisted of six separate runs. There were four runs of the phonetic categorization (PC) task, two for each continuum, (‘gift-kift’ and ‘giss-kiss’), and two runs for the tone categorization (TC) task. Eight of the final seventeen subjects received the runs in the order: gift-kift, giss-kiss, tone, gift-kift, giss-kiss, tone, and the remaining nine subjects received the runs in the order: giss-kiss, gift-kift, tone, giss-kiss, gift-kift, tone. Within each PC run, subjects heard ten repetitions of each of the seven points along the selected continuum, for a total of 70 stimuli per run, and 20 total presentations of each stimulus in the entire experiment. Each tone run consisted of ten repetitions of each of the two tones (High & Low) for 20 stimuli in each tone run, and 20 total repetitions of each tone stimulus. Stimuli within each run were presented in a fixed, pseudo-randomized order. Stimulus presentation and response registration were carried out using the same apparatus described in Experiment 1.

For the PC task, subjects were instructed to listen to each stimulus and to decide whether it started with a /g/ or a /k/ sound, indicating their choice by pressing one of two buttons using the right index finger. For the TC, subjects were instructed to listen to each tone and decide whether it was the High tone (1524 Hz) or the Low tone (457 Hz), and to press a corresponding button. Subjects were familiarized with all stimuli during the anatomical scan, and were allowed to practice 14 trials of the PC task and 10 trials of the TC task.

4.5.2.4. MR Imaging

Whole brain MRI was performed using the same apparatus and procedure as described in Experiment 2. As in Experiment 2, functional EPI images were acquired in a sparse sampling design, with each volume acquired in 1200 msec, followed by 800 msec of silence, yielding an effective volume repetition time of 2000 msec (see Figure 2) Auditory stimulus presentation was timed such that the stimulus fell in the center of the silent period. Stimulus types were distributed as evenly as possible into each of six trial onset asynchrony (TOA) bins (TOA= 2, 4, 6, 8, 10 or 12 sec) (see Appendix A for the distribution of stimuli to TOA bins). There was a four-volume offset between the start of EPI scanning and the onset of stimulus presentation in order to account for T1 saturation effects. These four volumes were discarded from subsequent analysis. Each of the four PC runs consisted of 247 EPI volumes, and each of the two TC runs consisted of 72 EPI volumes, for a total of 1132 acquired volumes.

4.5.2.5. Data Analysis

4.5.2.5.1. Behavioral analysis

Reaction time and categorization data were collected for all participants, with reaction times measured from the onset of the stimulus. Only responses with an RT of less than 2 seconds were allowed, so all presented data is effectively filtered to exclude responses greater than 2 seconds. Percent /g/ responses and RT for each subject were entered into separate two-way repeated measures ANOVAs, with VOT as one factor, and Continuum (‘gift’ or ‘giss’) as the other factor.

Categorization functions were plotted for each subject across the two continuum types, and the VOT boundary was calculated for each subject and continuum as described in the pilot experiment. All subjects’ boundary VOT values for each continuum were submitted to a one-way repeated measures ANOVA with ‘Boundary VOT’ as the dependent variable, and Continuum as the independent variable. Boundary values for each continuum were also compared to the boundary values from the nonword-nonword, gish—kish, continuum from Experiment 3A in individual one-way ANOVAs.   In addition, the stimulus number (1-7) which fell nearest to the calculated VOT boundary was determined for each continuum, and labeled as that subjects’ ‘boundary’ stimulus for that continuum.

4.5.2.5.2. MR analysis

4.5.2.5.2.1. Image Preprocessing

Image preprocessing was carried out as described in Experiment 1, with slice time correction carried out as described in Experiment 2.

4.5.2.5.2.2. Statistical Analysis

Deconvolution analyses were carried out as described in Experiment 1, with the exception that hemodynamic time courses were not extracted. Analysis was restricted to those voxels in which signal was recorded for all 17 subjects, which excluded voxels outside the brain (see Experiment 1). Voxel-wise percent signal change data for all 16 stimulus conditions (7 ‘gift’ stimuli + 7 ‘giss’ stimuli + 2 tone stimuli) were entered into a mixed factor ANOVA with subject as a random factor, and condition as a fixed factor. Planned comparisons included pairwise comparisons of each VOT stimulus with its partner in the opposing continuum (e.g. gift 4 – giss 4). The critical comparison in this study is between the boundary stimulus in one continuum against the same stimulus in the other continuum (i.e. gift 4 – giss 4, and gift 5 – giss 5). Additionally, to test whether stimuli identified as ‘boundary’ stimuli for a majority of subjects differed in activation pattern across the two continua, gift 5 was compared to giss 4. Finally, the effect of lexical status was examined by comparing the two continuum endpoints which were identified as words (gift 1 and giss 7) against the two continuum endpoints which were identified as non-words (gift 7 and giss 1).

Cluster thresholds were also determined as described in Experiment 1, using the same 17-subject clipping mask as used in the deconvolution analysis. All clustering was carried out at a corrected, cluster-level threshold of $p<0.05$ (voxel threshold of $p<0.025$, clusters consist of at least 62 contiguous voxels).

Anatomically defined regions of interest (ROIs) were defined based on the Talairach and Tournoux (1988) atlas, as implemented in AFNI. These regions were selected based on areas known to be modulated by changes in phonetic category structure (Experiment 1). They included the left IFG, MFG, insula, and angular gyrus, and the bilateral STG, Heschl’s gyri, and cingulate. Given that this analysis was largely conducted in order to confirm that the activation patterns varied as a function of phonetic category structure and were similar to the results observed in Experiment 1, stimuli were collapsed across continua in the following way. The phonetic category boundary stimulus was determined for each subject in each continuum, and those stimuli were assigned to the ‘boundary’ category. Next, stimuli which were one VOT step away from the boundary and in the ‘non-word’ direction on the continuum were grouped into the ‘Non-Word1’ category. Finally, stimuli which were one VOT step from the boundary in the ‘word’ direction on the continuum were grouped into the ‘Word1’ category, and those that were two steps in the word direction were grouped into the ‘Word2’ category. For instance, if for Subject A, gift5 and giss4 fell on the phonetic category boundary, these stimuli together constituted the ‘Boundary’ category for this subject. The stimuli which were one step towards the non-word end of the continuum for this subject were gift6 and giss3, which together would constitute the Non-Word1 category. The stimuli which were one step towards the word end of the continuum would be gift4 and giss5, which together made up the Word1 category, and the stimuli two steps towards the word end of the continuum were gift3 and giss6, and likewise constituted the Word2 category. It was not possible to create a Non-Word2 category for all subject, as 3 subjects had a boundary for ‘gift-kift’ which was only one step removed from the non-word end of the continuum.

Mean percent signal change was extracted from functionally and anatomically defined clusters for each subject and condition in order to quantitatively compare the activation patterns across conditions within a functionally defined region. These means were entered into a two way within-subjects, repeated measures ANOVA, with Continuum (gift or kiss) as one factor, and VOT as the other factor.

4.5.3. Results

4.5.3.1. Behavioral Results

Categorization data from subjects who showed lexical effects in the scanner (n=17) is shown in Figure 23. In general, subjects perceived stimuli categorically, with near-ceiling consistency in categorization of the first two and final two members of the continuum. A two-way ANOVA comparing the gift-kift and giss-kiss continua showed a significant main effect of Continuum (F(1, 16)=84.103, MSE=0.10, p<0.001), with more /g/ responses overall in the 'gift-kift' continuum, which reflects the shift in phonetic category boundary due to the lexical effect. There was also a significant main effect of VOT (F(6,96)=272.774, MSE=0.022, p<0.001), and a significant Continuum x VOT interaction (F(6,96)=21.526, MSE=0.10, p<0.001).

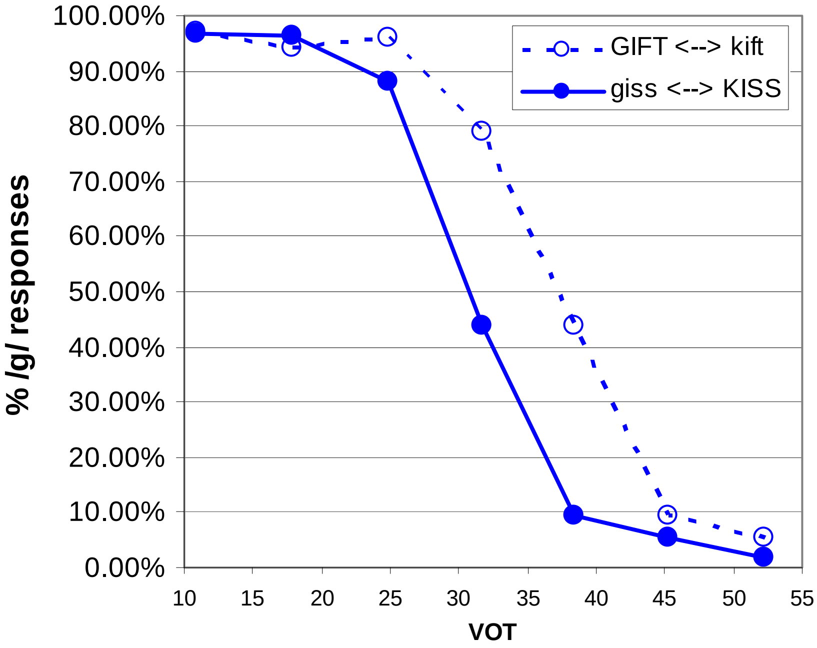

Figure 23. Mean percent /g/ responses for fMRI subjects (blue lines, n=17).

Reaction time data from these seventeen subjects is shown in Figure 24. As is typical of phonetic categorization experiments, subjects showed increased reaction times as stimuli approached the phonetic category boundary. In particular, subjects are the slowest to categorize stimuli which fall on the phonetic category boundary for a given continuum. This was reflected by a significant main effect of VOT stimulus (F(6,96)=15.432, MSE=3566.0, p<0.001). The two continua did not differ significantly in overall RT (F(1,16)=2.694, p>0.120), but there was a significant Continuum by VOT interaction (F(6,96)=14.176, MSE=2273.7, p<0.001). The reaction time peak for the gift—kift continuum was located at stimulus #5, and the RT peak for the giss—kiss continuum was located at stimulus #4.

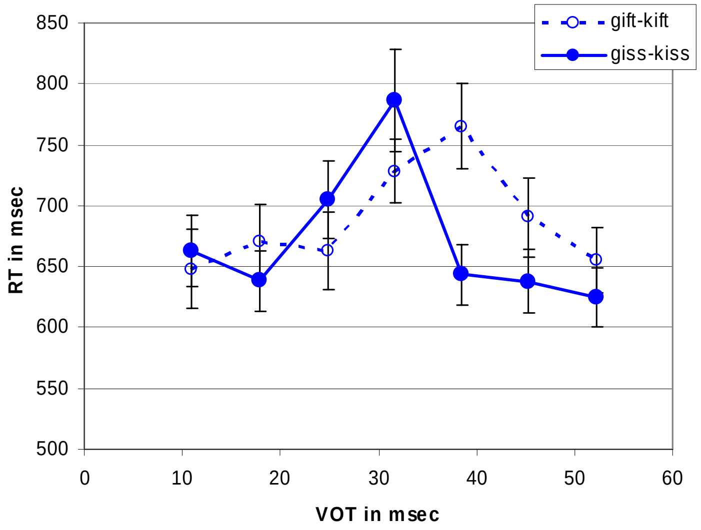

Figure 24. Reaction time data from 17 fMRI subjects.

Phonetic category boundary VOT values were calculated for each subject (see Table 10). The difference in boundary value between the two continua was statistically reliable (F(1,18)=52.550, MSE=6.261, p<0.001), indicating that subjects showed a significant lexical effect. When the boundary values for each continua were compared to the boundary value obtained in the lexically-unbiased Pilot 2, it was found that the boundary value for the gish-kish baseline was significantly different from the boundary for the gift-kift continuum, (F(1,28)=30.738, MSE=416.930, p<0.001), but was not significantly different from the giss-kiss boundary (F(1,28)=1.883, MSE=28.675, p<0.181). Hence, as in Pilot 1, only the gift-kift boundary shows a significant lexically-biased shift in identification function.

In order to ensure that each individual subject also showed a shift in phonetic category boundary between the two continua, the magnitude of the boundary shift ('giss-kiss' Boundary VOT – 'gift-kift' boundary VOT) was calculated for each subject (Table 10). As mentioned previously, in 17 of 19 subjects (all except cLE and qLE), the lexically-biased shift was large enough to result in a different stimulus identified as the boundary stimulus across the two continua. The mean size of the lexical effect was 5.89 msec (SD=3.54 msec). After eliminating the two subjects who did not show a lexical shift of at least one continuum step, the mean size of the lexical effect was 6.72 msec (SD=2.58). Most subjects (11 out of 17) had a category boundary which fell closest to stimulus #5 for the gift-kift, and most (again, 11 of 17 subjects) had a category boundary which was closest to stimulus #4 for the giss-kiss continuum. The distribution of boundary placement across the two continua is plotted in Figure 25.

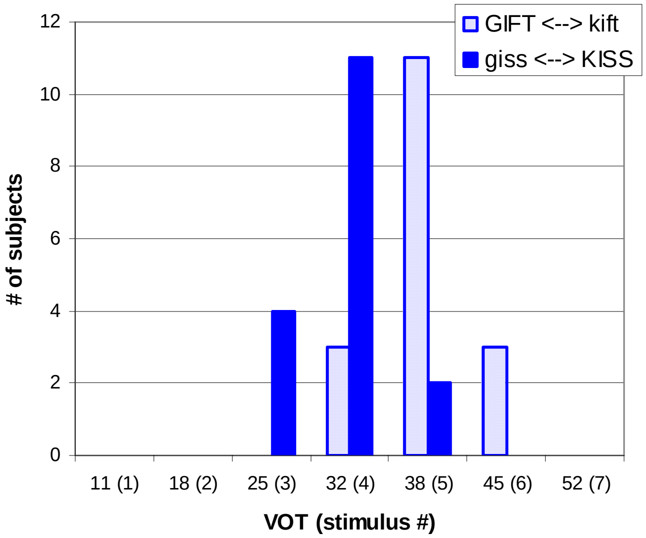

Figure 25. Histogram of phonetic category boundary placement for fMRI subjects (n=17).

Table 10. Phonetic category boundary values, listed in msec of VOT. Number in parentheses indicates the stimulus number (#1-#7) to which the boundary is nearest. The size of the lexical effect is calculated by subtracting the boundary value for 'giss-kiss' from the boundary value for 'gift-kift.' Subjects which did not show a lexical effect equal to at least one stimulus step are shown in bold.

Subject	*Boundary for gift-kift (closest stimulus value)*	*Boundary for giss-kiss (closest stimulus value)*	*Size of Lexical Effect (# of steps on continuum)*
aLE	35.07 (5)	31.27 (4)	3.80 (1)
bLE	32.86 (4)	25.86 (3)	7.00 (1)
cLE	**35.07 (5)**	**38.01 (5)**	**-2.93 (0)**
dLE	39.66 (5)	33.64 (4)	6.02 (1)
eLE	37.70 (5)	34.48 (4)	3.22 (1)
fLE	37.98 (5)	28.92 (4)	9.05 (1)
gLE	32.57 (4)	22.59 (3)	9.98 (1)
hLE	35.31 (5)	33.04 (4)	2.26 (1)
iLE	31.55 (4)	26.77 (3)	4.77 (1)
jLE	38.88 (5)	31.98 (4)	6.90 (1)
kLE	41.36 (5)	31.61 (4)	9.75 (1)
lLE	43.41 (6)	34.57 (4)	8.84 (2)
mLE	39.16 (5)	31.49 (4)	7.67 (1)
nLE	43.22 (6)	40.03 (5)	3.18 (1)
oLE	44.50 (6)	38.59 (5)	5.91 (1)
pLE	38.17 (5)	28.41 (4)	9.76 (1)
qLE	**29.12 (4)**	**28.67 (4)**	**0.45 (0)**
rLE	41.19 (5)	34.58 (4)	6.61 (1)
tLE	37.77 (5)	**28.20 (3)**	9.56 (2)
MEAN	**37.61 (5)**	**31.72 (4)**	**5.89 (1)**

4.5.3.2. Functional Results

4.5.3.2.1. Cluster analysis

A summary of all clusters activated in the planned comparisons which were significant at a corrected $p<0.05$ is shown in Table 11. The following discussion separates these comparisons into four categories. First, discussion focuses on the comparisons of particular interest with regards to the lexical effect phenomenon, that is, those that compare a boundary stimulus in one continuum to the same VOT value, non-boundary stimulus in the other continuum, that is, the Giss Boundary condition (gift4-giss4) and the Gift Boundary condition (gift5-gift5). Next, all other pairwise comparisons of the same VOT values across continua are reported, followed by a comparison designed to investigate potential differences in activation between words and non-words.

4.5.3.2.1.1. Cross-continuum comparisons involving boundary stimuli

Significant areas of activation were observed for both the Gift Boundary (gift5 - giss5) and Giss Boundary (gift4 -giss4) comparisons. In both comparisons, all clusters showed greater activation for the boundary stimulus (gift5 and giss 4, respectively) than the non-boundary stimulus. Gift5 - giss5 showed far greater differences in activation, both in the number of activated voxels, the extent of activation of those clusters, and the anatomical distribution of clusters. In fact, more total voxels were activated in this comparison (1889) than in any other comparison across VOT values. A cluster was found in the left STG which extended into Heschl’s gyrus and ventrally into the superior temporal sulcus (Figure 26, left). A similar cluster was found in the right hemisphere, centered in the STG, and extending medially into the insula, and into the right claustrum (Figure 26, right). The mean activation graphs for both clusters are also shown in Figure 26. Note that, in both cases, no difference in activation within these areas was seen in the gift4-giss4 comparison, in which there was no shift in the boundary owing to the lexical effect. In contrast, there was a difference in activation patterns in stimulus 5 indicating that superior temporal activation emerges for those boundary stimuli which have been influenced by the lexical effect. In general, increases in activation are expected in the superior temporal gyri as stimuli approach the phonetic category boundary (Experiments 1 & 2, Blumstein et al., 2005)). Of interest, STG activation emerges in these areas even when the onsets of the stimulus, and hence the placement of the stimulus along the phonetic continuum, are the same. The critical difference between gift5 and giss5 is that, due to lexical factors, one falls on the boundary and one does not. Thus, top-down lexical effects are seen to bias activation patterns in the superior temporal gyri, even for the same VOT value.

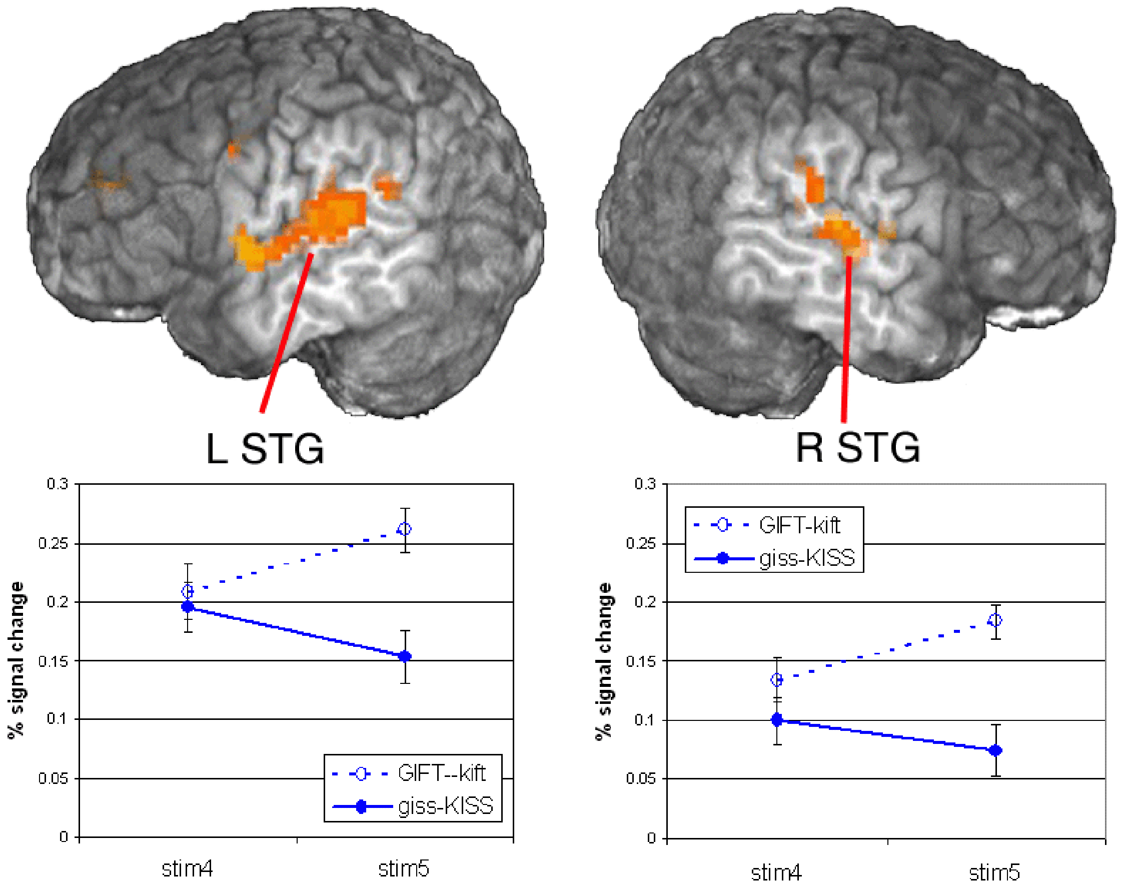

Figure 26. Clusters significant at a corrected threshold of $p<0.05$ ($p<0.025$, uncorrected, 62 contiguous voxels) for the gift5 – giss 5 comparison. All clusters are more active for gift5 (Boundary stimulus) than giss 5 (Non-Boundary stimulus). Activation displayed in percent signal change units. Left panel (x=-58) shows a cluster in the left STG (374 active voxels). Right panel (x=60) shows a cluster in the right STG (316 active voxels). Graphs at bottom indicate the mean percent signal change within these clusters for stimuli 4 and 5 within both gift-kift and giss-kiss continua.

In addition to the STG clusters, a very large cluster (960 voxels) was found in the left and right cingulate and anterior cingulate, which extended through left anterior white matter into the left insula (Figure 27A). Frontal clusters were observed in the left precentral gyrus and insula, and again further dorsally in the left MFG and superior frontal gyrus (Figure 27B&C). Finally, one midline cluster was observed in the precuneus.[4]

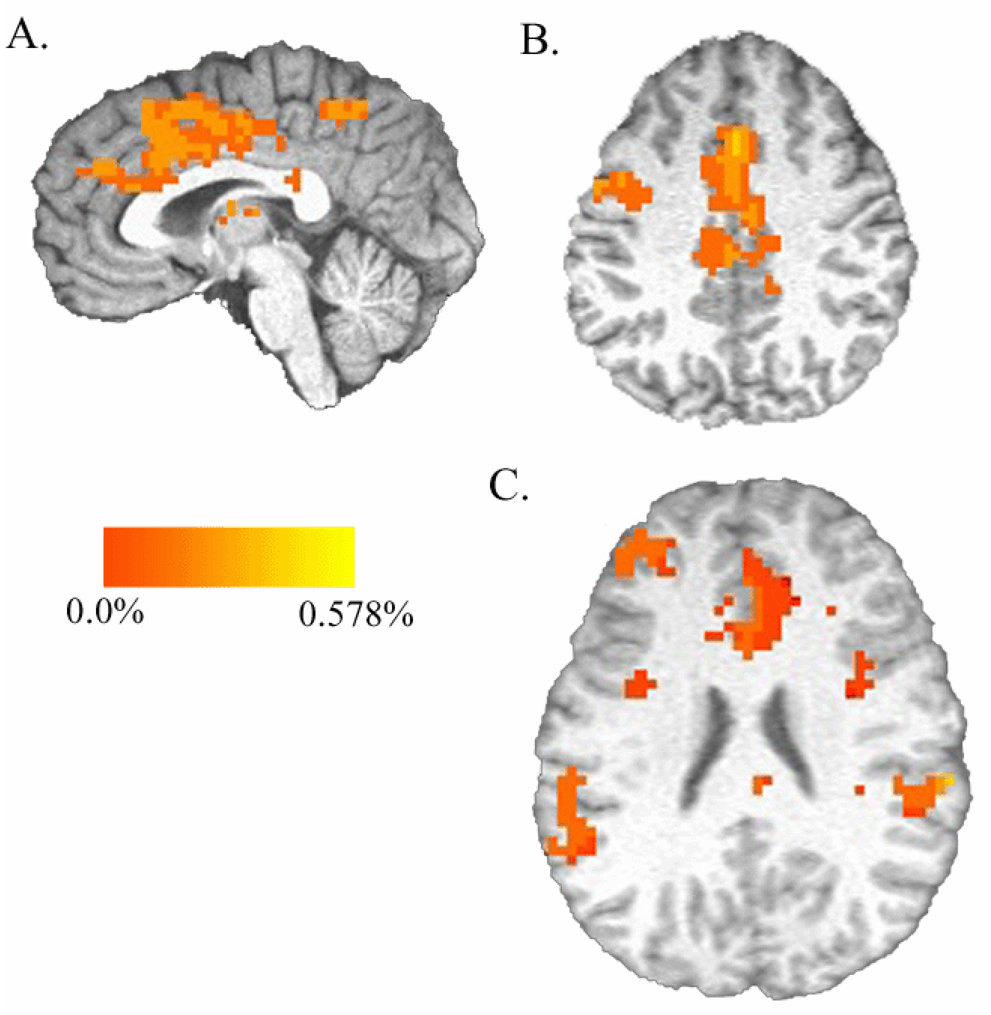

Figure 27. Clusters significant at a corrected threshold of $p<0.05$ ($p<0.025$, uncorrected, 62 contiguous voxels) for the gift5 – giss 5 comparison. All clusters are more active for gift5 (Boundary stimulus) than giss 5 (Non-Boundary stimulus). Activation displayed in percent signal change units. A. Sagittal slice at x=1, showing a large cluster centered on the cingulate gyrus. B. Axial slice at z=39, showing a left frontal cluster centered in the left precentral gyrus. Also visible is the large midline cluster shown in (A) C. Axial slice at z=24, showing a left frontal cluster centered in the left MFG. Also shows portions of midline and superior temporal clusters.

In contrast, the gift4-giss4 comparison yielded only two clusters (Figure 28). These clusters were in the bilateral inferior parietal lobules, with the right hemisphere

---

[4] One cluster was found in this analysis which was centered on a left ventricular area and encompassed some of the surrounding white matter. It was determined to be an artifact, and is not included in Table 10.

cluster extending somewhat into the superior parietal lobule. Significantly, no differences in activation were found between the gift4 (non-boundary) and giss 4 (boundary) stimuli in any of the areas previously shown to be sensitive to phonetic category structure, including left frontal areas, the STG areas, and midline areas.

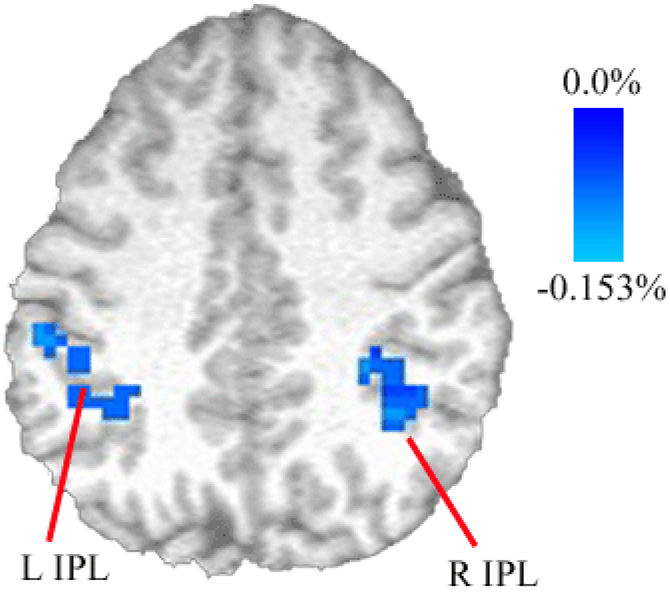

Figure 28. Clusters significant at a corrected threshold of $p<0.05$ ($p<0.025$, 62 contiguous voxels) for the gift4 – giss 4 comparison. All clusters more active for giss4 (Boundary stimulus) than giss4 (Non-boundary stimulus). Axial slice shown at z=40, depicting clusters in the left and right IPL.

In order to test whether there were inherent differences between stimuli which fell on the category boundary for most participants, gift5 was compared to giss4. The clusters activated for this comparison were very similar to those activated for the gift5 - giss5 comparison, suggesting that the gift boundary stimulus (gift5) shows fundamentally different patterns of activation from other stimuli (Table 10). Specifically, clusters emerged in the left and right superior temporal gyri which were similar in placement, although smaller in size to those activated in the gift5 - giss5 comparison. Additionally, a cluster in the left middle frontal gyrus was activated. Of interest, no cingulate activation emerged in this analysis, suggesting that the large cingulate cluster activated in the gift5 – giss5 comparison was also activated in the giss4 condition.

4.5.3.2.1.2. All other cross-continuum comparisons

Of the remaining cross-continuum statistical comparisons, one is notable in its similarity to the gift5-giss5 comparison. Like this comparison, the gift6-giss6 comparison yielded significant clusters in the superior temporal gyri bilaterally which were more active for the gift6 stimulus than the giss6 stimulus (Table 10). Of note, the gift6 – giss 6 comparison is the only other point along the VOT continuum where subjects categorized the gift continuum token as a boundary token (3 out of 17 subjects, see Figure 25), but where the giss continuum token does not fall on the boundary. Thus, it appears that, for a given point on the VOT continuum, if it is categorized as a boundary token in the gift continuum at all, but not as a boundary token in the giss continuum, modulation of activation in the superior temporal gyri emerges. As noted previously, when the giss continuum stimulus falls on the boundary, as in gift4 – giss 4 (for 11 of 17 subjects, giss4 falls on the boundary, Figure 25), no modulation of activation emerges in the superior temporal gyri. The same is true for gift3 – giss3 (for 4 of 17 subjects, giss3 falls on the boundary, Figure 25).

Greater activation in midline areas for stimuli in the gift continuum also emerges for several comparisons. Posterior cingulate and precuneus activation is seen for gift stimuli in the gift2 – giss2 comparison and the gift3 – giss3 comparison. Modulation of activation in the anterior cingulate and medial frontal gyri is also seen for stimuli in the gift3 – giss3 comparison. While midline activation is commonly reported in a variety of language tasks (Blumstein et al., 2005; Fiez et al., 1995; Jonides et al., 1998; Rissman et al., 2003), it is not clear what it signifies in the current experiment. If activation in this area were driven by the lexical status (i.e. word or non-word) of the stimulus, similar activation would be expected for the gift6 – giss6 or gift7 – giss7 comparisons, which is not the case. Activation in midline areas also typically emerges as decisions get more difficult, or as response conflict increases. As Figures C and D show, however, no significant differences emerge between the continua in either the reaction time or categorization of responses for gift2-giss2 or gift3-giss3.

A number of cross-continuum comparisons reveal modulation of activation in frontal areas. Aside from activations in frontal areas associated with the boundary comparisons (see above), no clear pattern emerges in terms of which conditions yield cross-continuum differences in frontal activation. The left IFG shows greater activation for giss1 in the gift1 – giss1 comparison. The left MFG shows greater activation for gift3 in the gift3 – giss3 comparison and for giss7 in the gift7 – giss7 comparison. There seem to be no trends with respect to the lexical status of the word—that is, a non-word (giss1) shows greater activation in frontal areas for the gift1 – giss1 comparison, while a word (giss7=kiss) shows greater activation in both left and right frontal areas in the gift7 – giss7 comparison.

While non-boundary, cross-continuum patterns of activation in frontal and midline areas do not adhere to any consistent pattern, the within-continuum pattern of activation in these areas will be shown to replicate the findings of Experiment 1 (see the region of interest analysis, 1.5.3.2.2).

4.5.3.2.1.3. Effects of lexical status

A potential confound in the design of this experiment is that, for any cross-continuum comparison of stimuli near the phonetic category boundary, one stimulus (the non-boundary stimulus) will be categorized such that it forms a real word, whereas the boundary stimulus will be equivocal in its lexical status. For instance, in the Gift Boundary comparison, gift5, which falls on the phonetic boundary, is equivocal in its lexical status--about half the time, subjects categorize it as starting with a /g/ sound, yielding ‘gift’, and half the time, with a /k/ sound, resulting in ‘kift’. Its VOT partner, giss5, however, is consistently categorized as beginning with a /k/ sound, and is hence heard consistently as ‘kiss.’ It is important to confirm that any differences observed when comparing gift5 and giss5 are due to processing differences inherent to phonetic category structure and not due to word vs. non-word differences in general. To this end, a statistical comparison was designed which contrasted continuum endpoints which are unequivocally real words (gift1 and giss7) against those that are unequivocally non-words (gift7 and giss1). No clusters emerged in this comparison, suggesting that any differences which emerged in the previous analyses were not due to lexical status per se of the stimuli.

Table 11. Areas of activation significant in planned comparisons, thresholded at a voxel-level threshold of p<0.025, cluster-level threshold of p<0.05 (62 contiguous voxels). Coordinates indicate the maximum intensity voxel for that cluster. All coordinates are in Talairach and Tournoux space.

	Maximum Intensity			*Number of activated voxels*	*Local Maximum*
Anatomical Region	*x*	*y*	*z*		
gift1 vs. giss1					
L IFG, L Precentral	-56	5	33	150	-0.1373
L. Subcortical: Thalamus, Caudate	-8	2	9	114	-0.1424
gift2 vs. giss2					
L/R Post. Cingulate, L/R Precuneus, LMTG	-5	-44	15	626	0.2279
Right Caudate, deep to Insula	14	17	-4	149	-0.1567
L. Precuneus	-5	-68	48	67	0.1762
gift3 vs. giss3					
L/R Post. Cingulate, L/R Precuneus	-2	-47	18	299	0.2922
L/R Medial frontal, L/R Ant. cingulate	-2	59	9	157	0.2471
L/R Ant Cingulate	-2	38	-4	76	0.1959
L MFG	-38	8	51	72	0.1298
gift4 vs. giss4					
LIPL	-47	-32	54	128	-0.1328
RSPL, RIPL	35	-68	48	83	-0.1529
gift5 vs. giss5					
L Cingulate, L/R Ant. Cingulate, R Insula	-5	11	36	960	0.1932
L STG, LTT, LSTP	-62	-8	6	374	0.2409
RSTG, R Insula, R Claustrum	62	-23	15	316	0.2738
L Precentral, Insula	-53	-2	42	85	0.1325
L MFG, L SFG	-29	50	30	78	0.1941
L/R Precuneus	-2	-56	51	76	0.1699
gift6 vs. giss6					
R STS, R STG,R MTG	62	-23	-4	193	0.2715
L STG	-47	-32	6	118	0.1158
R IFG, R MFG	53	8	18	74	0.1524
L Caudate, ventricle	-17	11	-1	63	0.1239
gift7 vs. giss7					
L MFG L Sup Frontal	-23	20	54	144	-0.2158
R MidOccipital, R MTG, portion of R STG	35	-80	33	126	-0.1904
R Sup Frontal , R MFG	20	14	54	78	-0.1607
R Precentral	47	-5	45	75	-0.1190
gift5 vs. giss4 (Gift Boundary vs. Giss Boundary)					
R STG, R MTG, R STS, R Insula, R STP	62	-29	-1	138	0.2272
L STG L Insula	-38	-26	12	89	0.1297
L MFG, L Sup Frontal	-29	38	24	73	0.1179
R STP, R Insula, R STG	62	-23	18	73	0.2435
WORD (endpoints) vs. NONWORD (endpoints)					
-no clusters found-					

4.5.3.2.2. ROI analysis

In order to confirm that the general within-continuum pattern of activation results was similar in this study to Experiments 1 and 2, an ROI analysis was carried out. In particular, it is expected that frontal areas such as the inferior frontal gyrus and anterior cingulate as well as posterior areas such as the bilateral STG should show graded activation, with greater activation as stimuli approach the phonetic boundary. This pattern is expected to emerge in frontal areas due to the executive demands of the phonetic categorization task, while activation in posterior areas might be presumed to be a result of the goodness of fit of a token to its phonetic category (see Experiments 1 and 2). In this analysis, each subject's fMRI data was aligned to his or her phonetic category boundary for each continuum, and analyses were carried out collapsing across continua (see Methods: Data Analysis for details). ROIs were selected based on the results of Experiments 1 and 2, which identified regions that show modulation of activation as a function of phonetic category structure.

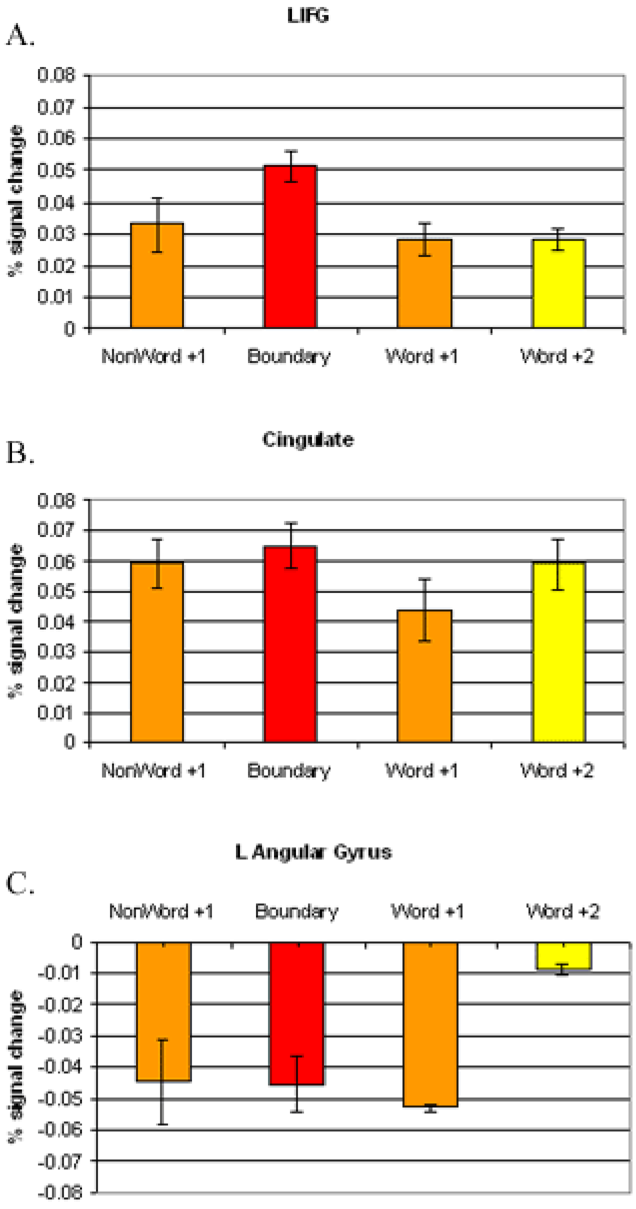

Figure 29. Mean percent signal change from anatomically defined regions of interest. Graphs shown for the left inferior frontal gyrus (A.), the bilateral cingulate gyri (B), and the left angular gyrus (C).

The left IFG showed a similar pattern of activation to that shown in Experiments 1 and 2, with the most activation for the stimuli which fell on the phonetic category boundary (Figure 29A). Unlike Experiment 1, however, the stimuli adjacent to the boundary (NW+1 and W+1) did not show increased activation compared to stimuli further from the boundary (W+2). This was confirmed by an ANOVA which revealed a main effect of continuum position (F(3,48)=3.731, MSE=0.001, p<0.017). Post-hoc tests revealed significant differences between the boundary stimuli and W+1 and W+2 stimuli, but no other significant difference between the boundary stimulus and the NW+1, and no significant differences within any of the non-boundary stimuli.

The left and right cingulate gyri also showed increased activation for the boundary stimulus compared to other stimuli (Figure 29B). However, results of an ANOVA showed that there was no main effect of continuum position on activation level (F(3,48)=1,266, MSE= p<0.001, p>0.297).

The pattern of results in the left angular gyrus also showed a pattern of results similar to that found in Experiment 1. In this area, all stimuli were deactivated with respect to the baseline (Figure 29C), but this deactivation was modulated by the placement of the stimulus on the continuum, with boundary stimuli and those just adjacent to the boundary (NW1 and W1) showing the most deactivation, and those further from the boundary (W2), showing much less deactivation. Results of an ANOVA show a main effect of boundary position (F(3,48)=2.844, MSE=0.002, p<0.047). Post-hoc tests show significant differences between W2 stimuli and all other stimuli, but no differences within boundary stimuli and NW1 and W1 stimuli. This is consistent with the pattern of results seen in Experiment 1, where endpoint stimuli (two steps from the category boundary) showed much less deactivation than boundary stimuli, or stimuli just adjacent to the category boundary.

Unlike Experiments 1 and 2, no differences emerged between stimuli in the right and left STG.

4.5.4. Discussion

4.5.4.1. Behavioral

In general, subjects perceived both VOT continua in a categorical manner, and showed typical increases in RT as stimuli approached the phonetic category boundary. This is consistent both with the behavioral literature on categorical perception (Liberman et al., 1957; Liberman et al., 1961; Pisoni & Tash, 1974), and also with MRI-collected behavioral data in Experiments 1 and 2. This behavioral pattern provides confirmatory evidence that subjects categorize stimuli generated from natural speech in much the same way as they do synthetic stimuli (Experiments 1 & 2).

Seventeen of nineteen subjects showed shifts in the phonetic category boundary in a direction consistent with the lexical effect. Importantly, the size of this effect was relatively large (6.72 msec) compared with other studies in the literature (see Pitt & Samuel, 1993 for a review). Several potential reasons exist for the size of the lexical effect in this study. First, although stimuli were generated from natural speech, only VOT was manipulated, which is only one of several acoustic cues that contribute to the perception of the phonetic feature of voicing (Repp, 1984). Evidence exists that the lexical effect disappears when acoustic cues are manipulated to create stimuli which more closely resemble the attributes of voicing found in natural speech (Burton et al., 1989; McQueen, 1991). Secondly, although stimuli in this study were presented during the pauses between MR scans, the listening conditions in the scanner are not equivalent to laboratory listening conditions. In spite of using noise dampening headphones, there was some degree of noise inherent in the scanning environment, and such noise has been shown to exaggerate the size of the lexical effect (Burton & Blumstein, 1995). It is worthy of note that natural listening conditions are rarely ideal, and lexical information may be recruited in noisy situations to aid processing (e.g. Broadbent, 1967) .

Additionally, for these seventeen subjects, the size of the lexical effect was large enough that different VOT stimuli were identified as the ‘boundary’ stimulus across the two continua (Table 9), with the boundary falling on stimulus #4 for the giss-kiss continuum for 11 of 17 subjects, and falling on stimulus #5 for the gift-kift continuum for 11 of 17 subjects (Table 9, Figure 25). Reaction time data also supported the presence of a lexical effect, with reaction time peaks at 32 msec (stimulus #4) for the ‘giss-kiss’ continuum, and at 38 msec (stimulus #5) for the ‘giss-kiss’ continuum. The fact that boundary values for the two continua fell on different VOT values allowed for activation comparisons within the same VOT value between stimuli which fell on the boundary with those that did not.

4.5.4.2. Functional

4.5.4.2.1. The Lexical Effect in the Temporal Lobes

A number of statistical comparisons revealed more activation in the right and left superior temporal gyri for the stimulus, gift5, which fell on the phonetic category boundary in the lexically-shifted continuum, gift-kift. This activation emerged when controlling for the VOT of the stimulus (gift5-giss5), and when contrasting this boundary stimulus to a boundary stimulus which had not undergone a lexical shift (gift5–giss4). Of particular interest, such activation did not emerge for the boundary stimulus of the non-shifted, giss-kiss continuum when controlling for VOT (gift4 – giss4).

The finding that activation in the STG shows different patterns for the same VOT value depending on the lexically-biased shift in boundary is of primary importance for several reasons. First, it demonstrates that the superior temporal gyri, which are known to be influenced by speech stimulus properties in general and phonetic category structure specifically, are also influenced by a lexically-biased shift in categorization function. The involvement of the STG in the mediation of the lexical effect provides strong evidence that the lexical effect itself has at the minimum a perceptual component to it. Second, increased activation in the STG only emerges when lexical information is both available to influence perception, and is actually used to shift the phonetic boundary. The lack of difference in the STG between the boundary stimulus for the giss continuum and its VOT matched partner (giss4-gift4) suggests that when lexical factors are unable to influence the locus of the phonetic category boundary, as is the case in the giss—kiss continuum, the STG responds to the stimuli as boundary stimuli, but shows no boosting in activation from lexical factors.

In sum, the sensitivity of the STG bilaterally to the lexical effect indicates that they are influenced by higher-level linguistic information, in this case, information about the lexical status of the word. At least three possibilities exist for this activation pattern. The most probable explanation is that this activation arises as a function of involvement of the STG themselves in both lexical and phonetic processing. Several studies (Giraud & Price, 2001; Poeppel et al., 2004; Wise et al., 2001) have shown activation of the superior temporal gyri for word stimuli. If the neural substrate for phonetic categorization and for lexical processing is one and the same, it would not be surprising that effects of lexical status on phonetic categorization processes are seen in these areas.

The second possibility is that modulation of activation in the STG is due to feedback to the STG from separate lexical-semantic areas such as the middle temporal gyrus or angular gyrus (see Binder & Price, 2001 for review). However, due to the lack of demonstrated modulation of activation in either area in any cross-continuum comparison, this seems improbable. If feedback from these lexical-semantic areas drives the modulation of activation in the STG, at least some differences among stimuli should be seen in these areas.

A third possibility is that activation patterns in the STG are influenced by feedback to the STG from decision-making areas. However, given the relative insensitivity of the STG to phonetic decision difficulty (c.f. Binder et al., 2004; Blumstein et al., 2005), it seems unlikely the activation pattern in the STG is solely a result of feedback from executive areas.

Although activation emerged in the STG for the lexically-shifted boundary stimulus, it is of note that the ROI analysis of these areas did not show significant modulation of activation as stimuli approached the phonetic category boundary, as they did in Experiments 1 and 2. One reason for this may be the normalization procedure used to collapse across stimuli in the ROI analysis. For this analysis, each subjects' data was aligned to his or her calculated category boundary, and averaged across VOT values and continua (See section 4.5.2.5.2.2.) For instance, the 'Boundary' category includes whichever stimuli fell on the boundary for a given subject, and so depending on the subject, includes giss3, giss4, or giss5, and gift4, gift5, or gift6 (see Figure 25). Thus, this analysis disregards the intrinsic acoustic-phonetic properties of the stimuli in favor of normalizing across behavioral patterns. If the STG are, as we have proposed, tuned specifically to these intrinsic properties, a lack of difference between conditions may simply reflect blurring across VOT values that occurred as a result of the normalization process.

4.5.4.2.2. Modulation of activation in frontal and midline areas

Modulation of activation in left frontal areas and midline areas is demonstrated for both individual cross-continuum comparisons as well as in the more general ROI analysis. The most consistent pattern of results is that activation emerges in these areas as a function of proximity to the phonetic category boundary, with the greatest activation emerging for stimuli which fall on the boundary itself. This pattern is consistent with the results of Experiments 1 and 2, which together provide evidence that these areas are sensitive to competition between phonetic categories, but less so to the goodness of fit of the input to any particular phonetic category. Of interest, significant differences between conditions emerge in LIFG in the normalized ROI analysis. As mentioned previously (1.5.4.2.1), the ROI analysis involves normalization of subject data according to their behavioral results. The fact that significant differences emerged in frontal areas in an analysis which specifically normalizes for subject behavior and hence reflects their decision processes supports the demonstrated sensitivity of left inferior frontal areas to executive components of phonetic categorization tasks.

4.5.4.2.3. Modulation of activation in parietal areas

Modulation of activation emerged in left and right inferior parietal cortex for the comparison of the giss boundary stimulus to its VOT-matched control (gift4-giss4, see Figure 28). Activation in parietal cortex did not emerge for any other comparison. The role of the parietal lobes in this study is unclear. Parietal activation in similar areas to those in this study have been linked to control of visual and spatial attention (Corbetta, Miezin, Dobmeyer, Shulman, & Petersen, 1991; Hopfinger, Buonocore, & Mangun, 2000), but rarely to auditory attention (Benedict et al., 1998), although it has been linked to auditory spatial localization (Bushara et al., 1999; Griffiths et al., 1998). Greater attentional resources may have been allocated to the boundary stimuli in this comparison, but it is unclear why such attention would be qualitatively different from that seen in the gift boundary comparison.

Areas inferior to the clusters shown in this study have been implicated in lexical and semantic processing, showing activation for word retrieval (Giraud & Price, 2001; Warburton et al., 1996), verbal working memory (Ravizza, Delgado, Chein, Becker, & Fiez, 2004), and have been implicated in semantic retrieval in general (see Price, 1998 for review). However, unlike in the current study, such activations nearly always extend into the angular gyrus, supramarginal gyrus, or the temporal lobes.

4.6. General Discussion and Conclusion

The lexical effect is one of several psycholinguistic phenomena which have been argued to demonstrate immediate use of top-down information in online processing. Such effects have been shown in the influence of visual context in online sentence comprehension (e.g. Altmann & Kamide, 1999; Spivey, Tanenhaus, Eberhard, & Sedivy, 2002), and in the use of written semantic context in the online interpretation of verbs with ambiguous categorization frames (e.g. Hare, McRae, & Elman, 2003). However, a crucial difference between these studies and the current one is that in these studies, high level contextual information is available either prior to or simultaneous with the presentation of the target stimulus. For instance, Altman & Kamide (1999) presented subjects with visual scene which depicted a child, a cake, and several toys. When subjects heard ‘The boy will eat . . .’ eye movements were immediately directed to the only edible object in the scene—the cake. However, this visual information was made available just prior to the onset of the sentence, and hence was fully available to subjects when they reached the verb in the sentence. In contrast, although cues to the voicing of the phonetic identity of the initial segment such as VOT and vowel length are available midway through the stimulus, in most studies of the lexical effect, the biasing contextual information, that is, the lexical status of the stimulus, is only available *after* processing the entire stimulus. In this sense, the lexical effect can never be said to be a pure ‘perceptual’ effect. Minimally, some processing of the initial phoneme must occur for subjects to identify the stimulus as ‘gift’ and not, for instance ‘bift.’ This processing must take place before lexical information is even available to play a role in biasing judgments of phonetic category identity. Indeed, evidence exists that lexical effects disappear when subjects make a speeded decision to the target and respond before hearing the end of the word (Fox, 1984; Miller & Dexter, 1988). However, this is not to say that lexical information might not be used very early in the processing stream, or indeed as soon as it becomes available. It is likely that in everyday speech, either the semantic content of the sentence or the discourse context is constraining enough that some information about the likely identity of an ambiguous word is available before the word is even presented. In such a case, context has the opportunity to influence perception of an initial phoneme as it is processed rather than in a post-hoc way.

The situation is somewhat different in the current study. In this study, stimuli were blocked by continuum, and within each block, all stimuli shared the same offset (/-Ift/ or /-Is/). Therefore, as soon as a subject had heard one stimulus, he or she could reliably predict the offset of the rest of the stimuli in the set. Due to this expectancy, the lexical status of any given stimulus was determined as soon as the initial phoneme was processed. That is, in a set where all stimuli end in /-Ift/, one only needs to determine that the initial phoneme is a /g/ to know that the stimulus is a real word, ‘gift.’ Arguably, in blocked studies of the lexical effect, lexical bias may be operating on a set-wise basis, re-calibrating the phonetic category boundary for the set as a whole, rather than for each individual stimulus. One possibility for the pattern of activation seen in this study is that STG sensitivity to the phonetic category boundary is ‘tuned’ by the characteristics of the stimulus set. In this situation, each run consisted of one continuum type, so subjects could reset the phonetic category boundary based on biasing lexical information on a run-by-run basis. Some evidence exists that the STG may be “reorganized” as a function of experience with speech stimuli. For instance, Liebenthal and colleagues (Liebenthal, Binder, Piorkowski, & Remez, 2003) showed that training in extracting phonetic information from sine wave speech modulated activation in the STG.

Ultimately, the involvement of the STG in the lexical effect provides evidence that, whether on a token-wise or set-wise basis, lexical information influences the perception of ambiguous phonetic tokens in a direct way. This is consistent with models of word recognition such as TRACE (McClelland & Elman, 1986) which allow for immediate use of higher-level information as it becomes available for processing. What is striking in the pattern of results seen in this study is the economy of processing that activation patterns reflect. The results from the boundary value comparisons show that activation in the STG is boosted only when lexical information is both available to bias perception, and is actually used to effect a change in phonetic category boundary. This suggests that the STG are more than just rote acoustic-phonetic analyzers of the incoming speech stream. Some information about either the stimulus set or the specific lexical properties of each individual stimulus must be available to neural structures involved in phonetic processing.

It is biologically sound for the speech processing system to have some degree of flexibility in adjusting the phonetic category boundary. Perception of phonetic category structure can be influenced by not only by lexical information but also by frequency of occurrence (Connine et al., 1993), neighborhood density (Newman et al., 1997), and speech rate (Miller, 1981a, 1981b), as well as other factors. Any cognitive model of speech processing must be able to accommodate such effects on perception. Likewise, any neural model of speech processing must also be able to integrate multiple sources of information in making phonetic decisions. If the STG is universally sensitive to higher-level sources of information on phonetic category structure, the STG bilaterally should be demonstrated to show modulation of activity whenever any of the above biasing factors, including speech rate, neighborhood density, and lexical frequency shift the placement of the phonetic boundary. Further research is necessary to determine whether this is, in fact the case.

Data from aphasic populations also supports the view that the lexical effect is mediated by posterior structures. Blumstein et al. (1994) showed that Broca's aphasics, who have damage to left anterior structures, showed a larger lexical effect than normals, whereas Wernicke's aphasics, who have damage to posterior temporal and parietal structures, showed no lexical effect. The posterior left STG often implicated in extracting speech-relevant properties of the stimulus (Binder et al., 2000; Poeppel et al., 2004; Scott & Wise, 2004) lies in close proximity to areas that are commonly implicated in lexical and semantic processing. It stands to reason, therefore that such an area might be the neural substrate for the integration of phonetic and lexical information.

5. CHAPTER 5: CONCLUSION

In three experiments, the neural structures underlying phonetic categorization processes were examined using fMRI. In all studies, modulation of activation in several peri-sylvian and midline neural structures was seen as subjects performed a phonetic categorization task on stimuli which ranged along a voice onset time continuum. The pattern of results demonstrated that neural structures are sensitive to relatively small (7-10 msec) acoustic changes in this continuum, and that this sensitivity can be measured using fMRI. Modulation of activation emerged in distinct neural areas depending on the degree of competition between phonetic categories and the goodness of fit of individual stimuli to their phonetic category, and as a function of the 'best fit' to the category and lexical influences on the phonetic boundary. These results have implications for both neural models of language processing in the brain as well as cognitive models of language processing. In particular, results support a model in which posterior portions of the bilateral superior temporal gyri subserve the analysis of the acoustic-phonetic properties of speech. Activation in these areas demonstrates sensitivity not only to phonetic category structure, but also to the effect of lexical information on the phonetic category. Left, and potentially also right inferior frontal areas, in contrast, seem to be responsible for resolving competition between phonetic categories, and as such show more activation as stimuli approach the category boundary. This pattern of data supports those models of language processing which allow for the spread of graded activation between levels of processing, as well as those which allow for feedback between lexical and phonetic levels of processing.

5.1. Superior temporal areas and their role in phonetic categorization

The STG bilaterally are sensitive to the structure of phonetic categories in the listener's inventory. In particular, more activation is shown for stimuli which are poorer exemplars of their phonetic category, indicating that some amount of processing must be performed on these stimuli in order to map them onto a phonetic category. In Experiment 1, the results of an ROI analysis indicated that there was more activation in the STG bilaterally for the boundary value stimulus than for other stimuli. It was also observed that activation in the bilateral STG did not correlate with difficulty of the phonetic decision as measured by reaction time, but did correlate with a regressor designed to model the goodness of fit of stimuli to their phonetic category. It was hypothesized on the basis of these results that the STG were responsive to the goodness of fit of a stimulus to its phonetic category, but not necessarily to the difficulty of the phonetic categorization decision. The results of Experiment 2 extended these findings, showing that not only was there increased activation for stimuli near the phonetic category boundary, but there was also increased activation for stimuli which were poor exemplars of their phonetic category that did not fall near the boundary. Taken together, the results of Experiments 1 and 2 suggest the role of the STG is in mapping incoming acoustic information onto a phonetic category. This mapping is more difficult when stimuli are poorer exemplars of their phonetic category, and as such the STG bilaterally show increasing activation for these poorer exemplars.

Some evidence exists which suggests that involvement of the STG bilaterally in processing acoustic-phonetic properties of speech is not limited to phonetic categorization tasks. First, the STG are activated in nearly every neuroimaging study of language comprehension (Hickok & Poeppel, 2000; Poeppel, 1996a) regardless of the task. The task-independence of activation in this area strongly implicates these areas in normal receptive language processing. Second, both MEG and ERP studies show electrophysiological signatures thought to arise from the left temporal lobe which are sensitive to phonetic category structure in passive listening tasks (Phillips, 2001; Sharma & Dorman, 1999; Sharma & Dorman, 2000). . Finally, a study by Guenther et al (Guenther et al., 2004) has shown increasing activation in the STG bilaterally for poorer exemplars of the vowel /i/ in a passive listening task, adding further support to the idea that the STG are sensitive to within-category phonetic variation in passive listening.

While Experiments 1 and 2 examined the effect of VOT on the perception of phonetic category structure, the results of Experiment 3 provided evidence that the STG may be ‘tuned’ to the phonetic category boundary on the basis of higher-level information such as that from the lexicon. In particular, lexical status influenced STG responsiveness to goodness of fit. In this experiment, the same VOT value showed more activation in the STG when the stimulus fell on the phonetic boundary due to lexical status, than when the same stimulus was in a continuum in which was not shifted by lexical status. Minimally, the results of Experiment 3 provide evidence that STG activation in response to phonetic category structure may be influenced by lexical information. This may occur due to the sensitivity of the STG itself to lexical information or to feedback to the STG from separate lexical areas such as the MTG or angular gyrus. If this interpretation is correct, then other lexical-level factors which have been shown to shift the phonetic category boundary, such as lexical frequency (Connine et al., 1993), and neighborhood density (Newman et al., 1997) should have a similar effect on activation in the STG in response to phonetic categorization. What is less clear is whether other sources of information such as sentential and semantic context which have been shown to shift phonetic identification functions would modulate activation in the STG as well. For instance, Borsky et al (1998) showed that the phonetic category boundary for stimuli which ranged along a VOT continuum from ‘goat’ to ‘coat’ could be shifted by a sentential context which biased the interpretation of the target word towards either ‘goat’ or ‘coat.’ It would be interesting to determine whether semantic context affects the modulation of activation in the STG as a function of phonetic category structure in the same way that lexical information does. If, indeed, the STG are sensitive to any factor which influences the placement of the phonetic boundary in phonetic categorization tasks, modulation of activation in the STG should occur as a function of the context manipulation.

5.1.1. Location of activation within the STG

While most models of language processing posit that the superior temporal gyri bilaterally are the basic substrate for speech processing, they differ in which portion of the STG are supposed to be responsible for this function (Binder, Frost, Hammeke, Rao, & Cox, 1996; Binder & Price, 2001; Hickok & Poeppel, 2000; Scott & Johnsrude, 2003). Hickock & Poeppel and Scott & Johnsrude agree that posterior portions of the left STG should be most active for analysis of phonetic category information. Binder & Price, in contrast, posit that the medial portion of the STG just lateral to Heschl’s gyrus should have the same role. Results from the present set of experiments are somewhat equivocal with respect to these claims. Experiment 1 showed that, in an arbitrary division of the left STG into thirds along its anterior to posterior extent, significant modulation of activation was observed only for the posterior portion, which is consistent with the Hickock & Poeppel as well as the Scott and Johnsrude models. It is of note, however, that although only the posterior third showed significant modulation, the pattern of activation was similar in all areas. Experiment 2 showed a different pattern of results. Significant modulation of activation was observed in both anterior and medial portions of the left STG, but not the most posterior portion. Again, a similar pattern of activation was observed for all thirds of the left STG. In this study, significant clusters in the left and right STG were also defined for comparisons of the best exemplars in the set with those stimuli which were poorer examples of the phonetic category (extreme and near-boundary tokens). These clusters centered primarily in medio-lateral portions of the STG, extending in to the STS, consistent with the Binder & Price (2001) hypothesis. A very similar set of bilateral STG clusters was defined in Experiment 4 in a comparison of a near-boundary VOT token in lexically shifted and un-shifted continua. Like the clusters defined in Experiment 3, these clusters were fairly medial. In general, the STG clusters activated in Experiments 2 and 3 were fairly large, and included portions of both medial and posterior segments of the STG. As such, no specific claims can be made on the basis of current data about functional sub-regions within the STG, except the suggestion that anterior areas of the STG do not seem to be involved in phonetic categorization to the same extent as medial and posterior areas.

The anterior temporal lobes have been implicated in analysis of spectrotemporal cues associated with speech (Scott & Johnsrude, 2003), and activation in a similar area bilaterally has been shown to correlate with sound identification accuracy (Binder et al., 2004). It is important to note that this area has been less linked to mapping of the incoming speech stream to a phonetic category, and more linked to the processing of the acoustic properties of speech generally. In the current studies, anterior temporal areas were less consistently modulated by changes in phonetic category structure than were posterior and medial areas. As discussed in Chatper 2, one possible explanation for this is that significant signal dropout occurred in the lateral anterior temporal lobes, although this dropout was variable between studies. Another possibility is that although the anterior temporal lobes may be involved analyzing the spectro-temporal properties of speech (Scott & Johnsrude, 2003; Wise et al., 2001), the differences in the spectro-temporal properties of speech stimuli in these studies were so minimal so as to engage the analysis process equally.

5.1.2. Lateralization of the STG response

The three neural models of language processing that were examined in the introduction all predict leftwards lateralization of activation in the STG for tasks which require contact with linguistic representations, such as speech categories or words (Binder & Price, 2001; Hickok & Poeppel, 2000; Scott & Johnsrude, 2003). Interestingly, in all three studies, there was significant bilateral involvement of the STG in processing phonetic category structure. In fact, in both Experiments 1 and 2 there was a suggestion that the right STG might be involved to a greater degree than the left STG. In Experiment 1, the ROI analysis of the right and left STG showed a greater activation for boundary tokens compared to other tokens in the right STG than the left. In Experiment 2, bilateral STG clusters were defined for both Extreme vs. Exemplar and Boundary vs. Exemplar comparisons. The right hemisphere clusters were respectively 53% and 68% larger than the left hemisphere clusters. Both of these studies used the same phonetic categorization task which presumably involves explicit access to the phonetic category. Nonetheless, no leftwards lateralization was seen in the STG, and evidence suggests that activation might instead be right-lateralized. Only in Experiment 3, which involved the influence of lexical information on phonetic categorization, was there any leftwards lateralization in the pattern of activation observed in the STG. In this study, the left hemisphere cluster was 18% larger than its counterpart in the right hemisphere. While this is not a particularly strong pattern of leftwards lateralization, taken together with the results of Experiments 1 and 2, it suggests that leftwards lateralization in the STG emerges only when information from lexical, and perhaps also from semantic levels of processing influences categorization processes.

5.2. Inferior frontal areas and their role in mediating competition

Results of Experiment 1 demonstrated that anterior structures (LIFG, anterior cingulate) showed graded activation, with increasing activation as stimuli approached the phonetic category boundary. In this study, activation in the left IFG was shown to correlate significantly with difficulty of the phonetic categorization decision (as measured by reaction time) but did not correlate with a “goodness of fit” regressor. A similar pattern of activation was also observed in Experiment 3, where each subject’s individual categorization data was aligned to his or her phonetic category boundary. Using this normalization process, significantly more activation was observed for stimuli which fell on the category boundary than for other stimuli. Experiment 2 also showed increasing activation as stimuli approached the phonetic category boundary, but also showed that activation did not increase in this area as stimuli became more difficult to process as a function of their poor fit to the phonetic category. Therefore, the results of Experiment 2 challenge the idea that the left inferior frontal gyrus is always active in proportion to task difficulty, while demonstrating that frontal areas are maximally sensitive to situations where there is competition between phonetic categories. Together these results are consistent with the view that the inferior frontal gyri, and in particular the left inferior frontal gyrus has a role resolving competition between linguistic alternatives (Thompson-Schill et al., 1997), rather than responsiveness to the stimulus goodness of fit.

The fact that increasing activation was not seen in this area as a function of the goodness of fit of exemplars to their phonetic category challenges the view that the left IFG is involved in resolving fine detail of the phonetic category, combined with an explicit judgment of phonetic category status (c.f. Zatorre et al., 1992; Zatorre et al., 1996). These authors propose that such fine examination of the stimulus is particularly needed when an explicit judgment of phonetic category status is required. Additionally, the stimuli used in this study are assumed to be approximately equivalent in terms of their segmentation demands. Given that this area was certainly activated for all speech stimuli in the phonetic categorization task, it is possible that the left IFG is recruited for a tasks which require segmentation of the initial phoneme from its phonetic context. However, the segmentation account of left IFG function would not predict the modulation of activation as a function of phonetic category structure observed in the left IFG, and as such cannot be a complete account of the function of this area (Burton, 2001; Burton et al., 2000).

Several models of language processing have posited a role for inferior frontal areas in accessing articulatory representations of phonetic category information (Hickok & Poeppel, 2000; Scott & Johnsrude, 2003). Such a system would be required by a ‘gestural’ theory of speech perception, which posits that the route to the perception of a phonetic category is via access to or modeling of the articulatory gestures used to produce that speech sound (Liberman & Mattingly, 1985). The motivation for the role of the IFG in mediating contact between perception and production systems stems in part from the historical view that patients with left IFG lesions often have speech production deficits, and may also show abnormal performance on phonetic categorization and discrimination tasks (Blumstein, Baker et al., 1977). It is proposed by Hickock, Poeppel, and colleagues that tasks such as the phonetic categorization task require explicit access to the phonetic categories of speech and hence must utilize a gestural mode of speech processing. In contrast, processes involved in auditory comprehension do not require contact with an articulatory representation. In their view, this hypothesis is supported by the general observation that inferior frontal areas are consistently activated in neuroimaging studies which require an explicit judgment about phonetic identity, but are not consistently activated when subjects passively listen to speech stimuli (Poeppel, 1996a).

The conclusion that inferior frontal areas are specifically involved in a gestural speech perception stream is not a necessary conclusion from the present series of studies, however. It is important to note that the left insula, which has been most consistently reported in lesion studies as relating to output deficits (Bates et al., 2003; Dronkers, 1996), and which has been implicated in speech production in neuroimaging studies (Wise, Greene, Buchel, & Scott, 1999) is not always activated in neuroimaging tasks which require attention to phonetic aspects of speech. Therefore the strongest version of the gestural stream claim, that the same areas are recruited for perception and production of speech, is not entirely supported. With regards to the data from the current series of studies, it is not clear how a gestural or articulatory theory might account for the variability observed in activation levels to tokens within a phonetic category. Presumably, tokens which fall near the phonetic category boundary are no harder to produce than those which are very far from the boundary, yet significant differences in activation were observed in frontal areas for the comparison between near-boundary and extreme tokens. In fact, some evidence exists to suggest that the near-boundary tokens used in this study (20 msec and 40 msec VOT) are actually produced more frequently than those from the extreme ends of the continuum (-100 and 160 msec VOT) (Lisker & Abramson, 1964). In sum, the proximity of areas such as the IFG to articulatory regions is no guarantee that they subserve a similar function.

A more general, executive role has also been suggested for the inferior frontal gyri. One hypothesis proposed is that these areas are involved in verbal working memory (Shivde & Thompson-Schill, 2004; Zarahn, Rakitin, Abela, Flynn, & Stern, 2004; Zurowski et al., 2002). The present pattern of results cannot be predicted by verbal working memory demands, as all trials were single stimuli, and did not need to be maintained in working memory. Another general executive account of inferior frontal function comes from Binder and colleagues. Binder et al. (Binder et al., 2004) propose that bilateral inferior frontal regions are involved in response selection during auditory tasks, based on the observation that activation in these areas was correlated with reaction time. The authors make no specific predictions, however, as to what conditions impose greater demands on response selection. At least one interpretation of their data, that is, that inferior frontal areas are active in proportion to response difficulty, is not supported by the present data. In Experiment 2, the stimuli on the extreme ends of the voicing continuum had slower reaction time latencies than the exemplar stimuli, yet no difference in activation was observed in inferior frontal areas between these two types of stimuli. The response selection claim, although more general in scope, is similar in nature to the claim made by Thompson-Schill and colleagues that left inferior frontal areas are involved in selection between competing alternatives (Thompson-Schill et al., 1997). It is this view which best explains the data obtained from the current series of studies.

As mentioned previously, competition may occur at several levels during the phonetic categorization task. In models of language comprehension such as TRACE (Gaskell & Marslen Wilson, 1999; McClelland & Elman, 1986), competition between activated representations is inherent in each level of processing. In the context of speech stimuli within a phonetic category, this means that tokens which are more similar to each other acoustically will compete with each other to a greater degree, even in normal language comprehension. It is also the case that the explicit nature of the phonetic categorization task imposes additional sources of competition to the task. In particular, competition may exist at the level of selection of a conceptual response ("is the token /da/ or /ta/?") or even at the level of executing the button press to indicate the selected item. With the current data, we are unable to determine which of these sources of competition contributes to the activation patterns observed in the IFG. These sources of competition could be disentangled by collecting data from participants who passively listened to stimuli which ranged along a continuum like the one presented in Experiment 2. If, in the absence of an overt response relating to phonetic category structure, participants still show competition-modulated activation for stimuli near the phonetic category boundary, this would lend support to the notion that competition at the level of the phonetic category itself is mediated by frontal structures.

5.3. The anterior cingulate gyrus and response selection

In all of the experiments reported here, greater activation was observed in the cingulate gyrus for tokens which fell on or near the phonetic category boundary than for other tokens. Like inferior frontal structures, the cingulate gyrus has been implicated in aspects of cognitive control. The cingulate is particularly implicated when there is competition at the level of the response (Milham et al., 2001; van Veen et al., 2001). As such, the increased activation for tokens which undergo larger amounts of competition (that is, those near or on the category boundary) is likely competition between the /da/ and /ta/ responses, rather than competition at the phonetic level of processing. As with the inferior frontal gyri, it is important to determine which aspects of activation in this area are dependent on the presence of an overt response. Given that this area is specifically implicated in response selection, it is unlikely that modulation of activation would be seen in this area for a task which did not require selection of an explicit response.

5.4. Implications for models of language processing

5.4.1. Graded activation in neural systems and language processing

The results of Experiment 1 demonstrated that graded activation could be observed for tokens within a phonetic category in a variety of neural systems, from those implicated in perceptual processing (STG), to those implicated in competition (IFG) and those involved in resolving response conflict (AC). Experiments 2 and 3 replicated this finding, showing within-category responsiveness in a variety of areas along the neural processing stream. To the extent to which it is believed that these brain areas also map onto separate components of phonetic processing, there is good evidence for graded spread of activation in the language system. Even the anterior cingulate, which is implicated in executive processing, but is not strictly a language area, shows graded sensitivity to tokens within a phonetic category. Therefore, some vestige of within-category information must be preserved as information is passed from perceptual/language areas to executive areas.

Graded activation was also observed in the IFG as a function the degree of competition that phonetic tokens undergo. The fact that a neural manifestation of competition exists lends some credence to those models of language processing which build in competitive influences within each level of processing (Gaskell & Marslen Wilson, 1997, 1999; McClelland & Elman, 1986; McNellis & Blumstein, 2001). Particular support of such a view would be given if competitive effects emerged in the IFG even under conditions which did not require overt categorization of the stimuli. This would suggest that effects of competition between phonetic categories are inherent in language comprehension generally, and are not specific to the phonetic categorization task.

5.4.2. The influence of the lexicon on perception

The modulation of activation observed in the bilateral STG which emerged as a function of the lexicality of the stimulus set indicates that, minimally speaking, there must be some influence at the perceptual level of lexical information on phonetic categorization. The STG seem to be very sensitive to phonetic category structure, and show modulation of activation when that structure is altered by the presence of biasing lexical items. As was discussed in Chapter 4, this pattern of results supports those models of language processing which allow for feedback from higher levels of processing to lower levels. It should be reiterated that the results of Experiment 3 do not indicate that there is *no* executive component to the lexical effect—indeed, the results of an ROI analysis in executive areas seems to suggest that there is a substantial executive component. It is most likely the case that the pattern of behavior observed in the lexical effect continua exists as the result of both perceptual and decision-related influences.

An interesting question is to what extent the phonetic category representations used in the STG are mutable, and what, if any aspects of these representations are fixed. Results of Experiment 3 showed that, for the same VOT value, there was more activation in the STG when the stimulus fell on the phonetic boundary due to lexical status, than when the same stimulus was in a continuum in which was not shifted by lexical status. Importantly, the STG did not show equal activation for all boundary stimuli; rather activation for the boundary value stimulus in the lexically-shifted continuum seemed to be boosted with respect to the boundary value stimulus in the un-shifted continuum (see Figure 26). The results of Experiment 3 seem to suggest that there is some latent response of the STG to the inherent structure of a listener's phonetic category, but that this response is influenced or 're-tuned' by non-phonetic factors such as lexical status. Some behavioral evidence exists to support the hypothesis that while phonetic category structure is relatively fixed and stable, it is subject to some influences from outside factors. For instance, Miller & colleagues (Miller & Dexter, 1988; Miller & Volaitis, 1989; Volaitis & Miller, 1992) have provided evidence that speech rate shifts the placement of the phonetic category boundary in phonetic categorization tasks. However, these shifts are not without limits: regardless of the speech rate, all phonetic category boundaries between voiced and voiceless stops fall between 20 and 45 msec VOT. Therefore, there may be some inherent restrictions to the mobility of the phonetic category boundary along the voicing continuum. Likewise, it may be the case that top-down influences in general tend to be a combination of a prepotent or dispositional bias in bottom-up processing, combined with information from higher levels of processing.

5.5. Conclusion

The results of this series of studies provide evidence for the sensitivity of neural structures and the fMRI signal to subtle variations in phonetic category structure. Perhaps more importantly, they also demonstrate a situation in which neuroimaging evidence may be brought to bear on debates which stem from the psycholinguistic rather than cognitive neuroscience literature. Experiments 1 and 2 provided evidence relevant to the question of whether graded activation spreads between levels of processing. Second, these two experiments also provided evidence of effects of competition between phonetic categories. Experiment 3 brought information to bear on the debate over whether lexical effects operate at a perceptual or decision-stage level of processing. While cross-disciplinary work between psycholinguistics and the cognitive neurosciences seems to have been embraced in the electrophysiological literature, such work is rarer in the fMRI and PET literature. It seems vital that as theories of the neural basis of language processing develop, there is contact with current psycholinguistic models of language. A more sophisticated view of language processes and mechanisms should provide cognitive neuroscientists with better testable hypotheses about the neural systems underlying these processes. Likewise, psycholinguists may increasingly find that neuroimaging offers a useful method for addressing questions which arise in that literature. Neuroimaging may be particularly powerful as a psycholinguistic tool when activation differences emerge between conditions which show no difference behaviorally. While a behavioral null result is often uninterpretable in the context of a reaction time study, activation differences for the same task conditions can reveal the operation of unexpected processes. In sum, a symbiotic relationship between psycholinguists and those studying the cognitive neuroscience of language can only help advance understanding in both fields.

References

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition, 73*(3), 247-264.

Andruski, J. E., Blumstein, S. E., & Burton, M. W. (1994). The effects of subphonetic differences on lexical access. *Cognition, 52*(3), 163-187.

Auerbach, S. H., Allard, T., Naeser, M., Alexander, M. P., & Albert, M. L. (1982). Pure word deafness. Analysis of a case with bilateral lesions and a defect at the prephonemic level. *Brain, 105*(Pt 2), 271-300.

Awh, E., Jonides, J., Smith, E. E., Schumacher, E. H., Koeppe, R. A., & Katz, S. (1996). Dissociation of storage and rehearsal in verbal working memory: Evidence from positron emission tomography. *Psychological Science, 7*(1), 25-31.

Baddeley, A. (1992). Working Memory. *Science*, *255*(5044), 556-559.

Basso, A., Casati, G., & Vignolo, L. A. (1977). Phonemic identification defect in aphasia. *Cortex, 13*(1), 85-95.

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

Belin, P., & Zatorre, R. J. (2000). Auditory perception and functional magnetic resonance imaging - fMRI. *Revue De Neuropsychologie, 10*(4), 603-621.

Belin, P., Zatorre, R. J., & Ahad, P. (2002). Human temporal-lobe response to vocal sounds. *Cognitive Brain Research, 13*(1), 17-26.

Belin, P., Zatorre, R. J., Hoge, R., Evans, A. C., & Pike, B. (1999). Event-related fMRI of the auditory cortex. *Neuroimage, 10*(4), 417-429.

Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature, 403*(6767), 309-312.

Benedict, R. H., Lockwood, A. H., Shucard, J. L., Shucard, D. W., Wack, D., & Murphy, B. W. (1998). Functional neuroimaging of attention in the auditory modality. *Neuroreport, 9*(1), 121-126.

Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S., Rao, S. M., & Cox, R. W. (1999). Conceptual processing during the conscious resting state. A functional MRI study. *J Cogn Neurosci, 11*(1), 80-95.

Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S. F., Springer, J. A., Kaufman, J. N., et al. (2000). Human temporal lobe activation by speech and nonspeech sounds. *Cerebral Cortex, 10*(5), 512-528.

Binder, J. R., Frost, J. A., Hammeke, T. A., Cox, R. W., Rao, S. M., & Prieto, T. (1997). Human brain language areas identified by functional magnetic resonance imaging. *Journal of Neuroscience, 17*(1), 353-362.

Binder, J. R., Frost, J. A., Hammeke, T. A., Rao, S. M., & Cox, R. W. (1996). Function of the left planum temporale in auditory and linguistic processing. *Brain, 119*, 1239-1247.

Binder, J. R., Liebenthal, E., Possing, E. T., Medler, D. A., & Ward, B. D. (2004). Neural correlates of sensory and decision processes in auditory object identification. *Nat Neurosci, 7*(3), 295-301.

Binder, J. R., & Price, C. (2001). Functional Neuroimaging of Language. In R. Cabeza & A. Kingstone (Eds.), *Handbook of Functional Neuroimaging of Cognition* (pp. 187-251). Cambridge: MIT Press.

Binder, J. R., Rao, S. M., Hammeke, T. A., Yetkin, F. Z., Jesmanowicz, A., Bandettini, P. A., et al. (1994). Functional magnetic resonance imaging of human auditory cortex. *Annals of Neurology, 35*(6), 662-672.

Blumstein, S. E. (1995). The Neurobiology of the Sound Structure of Language. In M. S. Gazzaniga (Ed.), *The Cognitive Neurosciences* (pp. 915-930). Cambridge: MIT Press.

Blumstein, S. E. (1998). Phonological Aspects of Aphasia. In M. T. Sarno (Ed.), *Acquired Aphasia* (3 ed., pp. 157-186). Boston: Academic Press.

Blumstein, S. E., Baker, E., & Goodglass, H. (1977). Phonological factors in auditory comprehension in aphasia. *Neuropsychologia, 15*(1), 19-30.

Blumstein, S. E., Burton, M. W., Baum, S., Waldstein, R., & Katz, D. (1994). The role of lexical status on the phonetic categorization of speech in aphasia. *Brain and Language*, *46*(2), 181-197.

Blumstein, S. E., Cooper, W. E., Zurif, E. B., & Caramazza, A. (1977). The perception and production of voice-onset time in aphasia. *Neuropsychologia, 15*(3), 371-383.

Blumstein, S. E., Myers, E. B., & Rissman, J. (2005). The perception of voice-onset time: An fMRI investigation of phonetic category structure. *Journal of Cognitive Neuroscience*.

Blumstein, S. E., Tartter, V. C., Nigro, G., & Statlender, S. (1984). Acoustic cues for the perception of place of articulation on aphasia. *Brain and Language*, *22*(1), 128-149.

Borsky, S., Tuller, B., & Shapiro, L. P. (1998). "How to milk a coat:" the effects of semantic and acoustic information on phoneme categorization. *J Acoust Soc Am, 103*(5 Pt 1), 2670-2676.

Brady, S. A., & Darwin, C. J. (1978). Range effect in the perception of voicing. *J Acoust Soc Am, 63*(5), 1556-1558.

Broadbent, D. E. (1967). Word-frequency effect and response bias. *Psychological Review, 74*(1), 1-15.

Buckner, R. L., & Logan, J. M. (2001). Functional Neuroimaging Methods: PET and fMRI. In R. Cabeza & A. Kingstone (Eds.), *Handbook of Functional Neuroimaging of Cognition* (pp. 187-251). Cambridge: MIT Press.

Burton, M. W. (2001). The role of inferior frontal cortex in phonological processing. *Cognitive Science, 25*(5), 695-709.

Burton, M. W., Baum, S. R., & Blumstein, S. E. (1989). Lexical effects on the phonetic categorization of speech: The role of acoustic structure. *Journal of Experimental Psychology: Human Perception and Performance, 15*(3), 567-575.

Burton, M. W., & Blumstein, S. E. (1995). Lexical effects on phonetic categorization: The role of stimulus naturalness and stimulus quality. *Journal of Experimental Psychology: Human Perception and Performance, 21*(5), 1230-1235.

Burton, M. W., Small, S. L., & Blumstein, S. E. (2000). The role of segmentation in phonological processing: An fMRI investigation. *Journal of Cognitive Neuroscience, 12*(4), 679-690.

Bushara, K. O., Weeks, R. A., Ishii, K., Catalan, M. J., Tian, B., Rauschecker, J. P., et al. (1999). Modality-specific frontal and parietal areas for auditory and visual spatial localization in humans. *Nat Neurosci, 2*(8), 759-766.

Caplan, D., Gow, D., & Makris, N. (1995). Analysis of lesions by MRI in stroke patients with acoustic-phonetic processing deficits. *Neurology, 45*(2), 293-298.

Carter, C. S., Braver, T. S., Barch, D. M., Botvinick, M. M., Noll, D., & Cohen, J. D. (1998). Anterior cingulate cortex, error detection, and the online monitoring of performance. *Science, 280*(5364), 747-749.

Cohen, M. S. (1997). Parametric analysis of fMRI data using linear systems methods. *Neuroimage, 6*(2), 93-103.

Coltheart, M. (1981). The MRC Psycholinguistic Database. *Quarterly Journal of Experimental Psychology, 33*(A), 497-505.

Connine, C. M., Titone, D., & Wang, J. (1993). Auditory word recognition: extrinsic and intrinsic effects of word frequency. *Journal of Experimental Psychology: Learning Memory and Cognition, 19*(1), 81-94.

Corbetta, M., Miezin, F. M., Dobmeyer, S., Shulman, G. L., & Petersen, S. E. (1991). Selective and divided attention during visual discriminations of shape, color, and speed: functional anatomy by positron emission tomography. *J Neurosci, 11*(8), 2383-2402.

Cox, R. W., & Hyde, J. S. (1997). Software tools for analysis and visualization of fMRI data. *NMR in Biomedicine, 10*(4-5), 171-178.

Cox, R. W., & Jesmanowicz, A. (1999). Real-time 3D image registration for functional MRI. *Magnetic Resonance in Medicine, 42*(6), 1014-1018.

Damasio, H. (1998). Neuroanatomical Correlates of Aphasia. In M. T. Sarno (Ed.), *Acquired Aphasia* (3rd ed., pp. 43-70). New York: Academic Press.

Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychol Rev, 104*(4), 801-838.

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.

Demonet, J. F., Price, C., Wise, R., & Frackowiak, R. S. J. (1994). A PET study of cognitive strategies in normal subjects during language tasks - Influence of phonetic ambiguity and sequence processing on phoneme monitoring. *Brain, 117*, 671-682.

Dronkers, N. F. (1996). A new brain region for coordinating speech articulation. *Nature, 384*(6605), 159-161.

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

Engelien, A., & Huber, W. (2001). SK - A case of pure word deafness due to unilateral left hemisphere lesion. *Brain and Language, 79*(1), 35-38.

Eustache, F., Lechevalier, B., Viader, F., & Lambert, J. (1990). Identification and discrimination disorders in auditory perception: a report on two cases. *Neuropsychologia, 28*(3), 257-270.

Fiez, J. A., Raichle, M. E., Miezin, F. M., Petersen, S. E., Tallal, P., & Katz, W. F. (1995). PET studies of auditory and phonological processing - Effects of stimulus characteristics and task demands. *Journal of Cognitive Neuroscience, 7*(3), 357-375.

Fox, R. A. (1984). Effect of lexical status on phonetic categorization. *Journal of Experimental Psychology: Human Perception and Performance, 10*(4), 526-540.

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance, 6*(1), 110-125.

Gaskell, M., & Marslen Wilson, W. D. (1997). Integrating form and meaning: A distributed model of speech perception. *Language and Cognitive Processes, 12*(5-6), 613-656.

Gaskell, M., & Marslen Wilson, W. D. (1999). Ambiguity, competition, and blending in spoken word recognition. *Cognitive Science, 23*(4), 439-462.

Gehring, W. J., & Knight, R. T. (2000). Prefrontal-cingulate interactions in action monitoring. *Nature Neuroscience, 3*(5), 516-520.

Geschwind, N. (1965). Disconnexion syndromes in animals and man. I. *Brain, 88*(2), 237-294.

Giraud, A. L., Lorenzi, C., Ashburner, J., Wable, J., Johnsrude, I., Frackowiak, R., et al. (2000). Representation of the temporal envelope of sounds in the human brain. *J Neurophysiol, 84*(3), 1588-1598.

Giraud, A. L., & Price, C. J. (2001). The constraints functional neuroimaging places on classical models of auditory word processing. *Journal of Cognitive Neuroscience, 13*(6), 754-765.

Goodglass, H. (1993). *Understanding aphasia*: San Diego, CA, US: Academic Press, Inc.

Griffiths, T. D., Rees, A., & Green, G. G. R. (1999). Disorders of human complex sound processing. *Neurocase, 5*(5), 365-378.

Griffiths, T. D., Rees, G., Rees, A., Green, G. G., Witton, C., Rowe, D., et al. (1998). Right parietal cortex is involved in the perception of sound movement in humans. *Nat Neurosci, 1*(1), 74-79.

Guenther, F. H., & Gjaja, M. N. (1996). The perceptual magnet effect as an emergent property of neural map formation. *J Acoust Soc Am, 100*(2 Pt 1), 1111-1121.

Guenther, F. H., Husain, F. T., Cohen, M. A., & Shinn-Cunningham, B. G. (1999). Effects of categorization and discrimination training on auditory perceptual space. *J Acoust Soc Am, 106*(5), 2900-2912.

Guenther, F. H., Nieto-Castanon, A., Ghosh, S. S., & Tourville, J. A. (2004). Representation of sound categories in auditory cortical maps. *J Speech Lang Hear Res, 47*(1), 46-57.

Gusnard, D. A., & Raichle, M. E. (2001). Searching for a baseline: functional imaging and the resting human brain. *Nat Rev Neurosci, 2*(10), 685-694.

Hall, D. A., Haggard, M. P., Akeroyd, M. A., Palmer, A. R., Summerfield, A. Q., Elliott, M. R., et al. (1999). "Sparse" temporal sampling in auditory fMRI. *Human Brain Mapping, 7*(3), 213-223.

Hare, M., McRae, K., & Elman, J. L. (2003). Sense and structure: Meaning as a determinant of verb subcategorization prefences. *Journal of Memory and Language, 48*(2), 281-303.

Hickok, G., & Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception. *Trends Cogn Sci, 4*(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*(1-2), 67-99.

Hopfinger, J. B., Buonocore, M. H., & Mangun, G. R. (2000). The neural mechanisms of top-down attentional control. *Nat Neurosci, 3*(3), 284-291.

Howard, D., Patterson, K., Wise, R., Brown, W. D., Friston, K., Weiller, C., et al. (1992). The cortical localization of the lexicons - Positron emission tomography evidence. *Brain, 115*, 1769-1782.

Hutchison, E., Blumstein, S. E., & Myers, E. B. (submitted). An event-related fMRI investigation of voice-onset time discrimination.

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*(29), 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.

Jonides, J., Schumacher, E. H., Smith, E. E., Koeppe, R. A., Awh, E., Reuter-Lorenz, P. A., et al. (1998). The role of parietal cortex in verbal working memory. *J Neurosci, 18*(13), 5026-5034.

Kan, I. P., & Thompson-Schill, S. L. (2004a). Effect of name agreement on prefrontal activity during overt and covert picture naming. *Cogn Affect Behav Neurosci, 4*(1), 43-57.

Kan, I. P., & Thompson-Schill, S. L. (2004b). Selection from perceptual and conceptual representations. *Cogn Affect Behav Neurosci, 4*(4), 466-482.

Keating, P. A., Mikos, M. J., & Ganong, W. F. (1981). A cross-language study of range of voice onset time in the perception of intial stop voicing. *J Acoust Soc Am, 70*(5), 1261-1271.

Kessinger, R. H., & Blumstein, S. E. (manuscript). The influence of phonetic category structure on lexical access.

Klatt, D. H. (1980). Software for a cascade/parallel formant synthesizer. *Journal of the Acoustical Society of America, 67*(3), 971-995.

Kucera, H., & Francis, W. N. (1967). *Computational Analysis of Present-Day American English*. Providence: Brown University Press.

Kuhl, P. K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Percept Psychophys, 50*(2), 93-107.

Lancaster, J. L., Woldorff, M. G., Parsons, L. M., Liotti, M., Freitas, C. S., Rainey, L., et al. (2000). Automated Talairach atlas labels for functional brain mapping. *Hum Brain Mapp, 10*(3), 120-131.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*.

Liberman, A. M., Harris, K. S., Kinney, J. A., & Lane, H. (1961). The discrimination of relative onset-time of the components of certain speech and nonspeech patterns. *Journal of Experimental Psychology, 61*, 379-388.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition, 21*(1), 1-36.

Liebenthal, E., Binder, J. R., Piorkowski, R. L., & Remez, R. E. (2003). Short-term reorganization of auditory analysis induced by phonetic experience. *J Cogn Neurosci, 15*(4), 549-558.

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: acoustical measurements. *Word, 20*, 384-422.

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: the neighborhood activation model. *Ear Hear, 19*(1), 1-36.

MacDonald, A. W., 3rd, Cohen, J. D., Stenger, V. A., & Carter, C. S. (2000). Dissociating the role of the dorsolateral prefrontal and anterior cingulate cortex in cognitive control. *Science, 288*(5472), 1835-1838.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18*(1), 1-86.

McKiernan, K. A., Kaufman, J. N., Kucera-Thompson, J., & Binder, J. R. (2003). A parametric manipulation of factors affecting task-induced deactivation in functional neuroimaging. *J Cogn Neurosci, 15*(3), 394-408.

McMurray, B. (in preparation). A [somewhat] new systematic approach to formant-based speech synthesis for empirical research.

McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition, 86*(2), B33-42.

McMurray, B., Tanenhaus, M. K., Aslin, R. N., & Spivey, M. J. (2003). Probabilistic constraint satisfaction at the lexical/phonetic interface: evidence for gradient effects of within-category VOT on lexical access. *J Psycholinguist Res, 32*(1), 77-97.

McNellis, M. G., & Blumstein, S. E. (2001). Self-organizing dynamics of lexical access in normals and aphasics. *J Cogn Neurosci, 13*(2), 151-170.

McQueen, J. M. (1991). The influence of the lexicon on phonetic categorization: Stimulus quality in word-final ambiguity. *Journal of Experimental Psychology: Human Perception and Performance, 17*(2), 433-443.

Mertus, J. (1989). BLISS user's manual. Providence: Brown University.

Miceli, G. (1982). The processing of speech sounds in a patient with cortical auditory disorder. *Neuropsychologia, 20*(1), 5-20.

Milham, M. P., Banich, M. T., Webb, A., Barad, V., Cohen, N. J., Wszalek, T., et al. (2001). The relative involvement of anterior cingulate and prefrontal cortex in attentional control depends on nature of conflict. *Brain Res Cogn Brain Res, 12*(3), 467-473.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annu Rev Neurosci, 24*, 167-202.

Miller, G. A., Heise, G. A., & Lichten, W. (1951). The intelligibility of speech as a function of the context of the test materials. *Journal of Experimental Psychology, 41*, 329-335.

Miller, J. L. (1981a). Effects of speaking rate on segmental distinctions. In P. D. Eimas & J. L. Miller (Eds.), *Perspectives on the study of speech* (pp. 39-74). Hillsdale, NJ: Erlbaum.

Miller, J. L. (1981b). Some effects of speaking rate on phonetic perception. *Phonetica, 38*(1-3), 159-180.

Miller, J. L., & Dexter, E. R. (1988). Effects of speaking rate and lexical status on phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance, 14*(3), 369-378.

Miller, J. L., Green, K. P., & Reeves, A. (1986). Speaking rate and segments: A look at the relation betwen speech production and speech perception for the voicing contrast. *Phonetica, 43*, 106-115.

Miller, J. L., & Volaitis, L. E. (1989). Effect of speaking rate on the perceptual structure of a phonetic category. *Percept Psychophys, 46*(6), 505-512.

Motomura, N., Yamadori, A., Mori, E., & Tamaru, F. (1986). Auditory agnosia. Analysis of a case with bilateral subcortical lesions. *Brain, 109*(Pt 3), 379-391.

Mummery, C. J., Ashburner, J., Scott, S. K., & Wise, R. J. S. (1999). Functional neuroimaging of speech perception in six normal and two aphasic subjects. *Journal of the Acoustical Society of America, 106*(1), 449-457.

Naatanen, R., & Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. *Psychophysiology, 24*(4), 375-425.

Newman, R. S., Sawusch, J. R., & Luce, P. A. (1997). Lexical neighborhood effects in phonetic processing. *Journal of Experimental Psychology: Human Perception and Performance, 23*(3), 873-889.

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition, 52*(3), 189-234.

Norris, D., McQueen, J. M., & Cutler, A. (1995). Competition and segmentation in spoken-word recognition. *J Exp Psychol Learn Mem Cogn*, *21*(5), 1209-1228.

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*, *9*(1), 97-113.

Paradis, A. L., Van de Moortele, P. F., Le Bihan, D., & Poline, J. B. (2001). Slice acquisition order and blood oxygenation level dependent frequency content: an event-related functional magnetic resonance imaging study. *Magma*, *13*(2), 91-100.

Paulesu, E., Frith, C. D., & Frackowiak, R. S. J. (1993). The neural correlates of the verbal component of working memory. *Nature, 362*(6418), 342-345.

Petersen, S. E., Fox, P. T., Posner, M. I., Mintun, M., & Raichle, M. E. (1988). Positron emission tomographic studies of the cortical anatomy of single-word processing. *Nature*, *331*(6157), 585-589.

Petersen, S. E., Fox, P. T., Posner, M. I., Mintun, M., & Raichle, M. E. (1989). Positron emission tomographic studies of the processing of single words. *Journal of Cognitive Neuroscience, 1*(2), 153-170.

Phillips, C. (2001). Levels of representation in the electrophysiology of speech perception. *Cognitive Science, 25*(5), 711-731.

Phillips, C., Pellathy, T., Marantz, A., Yellin, E., Wexler, K., Poeppel, D., et al. (2000). Auditory cortex accesses phonological categories: an MEG mismatch study. *J Cogn Neurosci, 12*(6), 1038-1055.

Pisoni, D. B., & Tash, J. (1974). Reaction times to comparisons within and across phonetic categories. *Perception and Psychophysics, 15*, 289-290.

Pitt, M. A. (1995). The locus of the lexical shift in phoneme identification. *Journal of Experimental Psychology: Learning Memory and Cognition, 21*(4), 1037-1052.

Pitt, M. A., & Samuel, A. G. (1993). An empirical and meta-analytic evaluation of the phoneme identification task. *Journal of Experimental Psychology: Human Perception and Performance, 19*(4), 699-725.

Poeppel, D. (1996a). A critical review of PET studies of phonological processing. *Brain and Language, 55*(3), 317-351.

Poeppel, D. (1996b). Some remaining questions about studying phonological processing with PET: Response to Demonet, Fiez, Paulesu, Petersen, and Zatorre (1996). *Brain and Language, 55*(3), 380-385.

Poeppel, D. (2001). Pure word deafness and the bilateral processing of the speech code. *Cognitive Science, 25*(5), 679-693.

Poeppel, D., Guillemin, A., Thompson, J., Fritz, J., Bavelier, D., & Braun, A. R. (2004). Auditory lexical decision, categorical perception, and FM direction discrimination differentially engage left and right auditory cortex. *Neuropsychologia, 42*(2), 183-200.

Poldrack, R. A., Temple, E., Protopapas, A., Nagarajan, S., Tallal, P., Merzenich, M., et al. (2001). Relations between the neural bases of dynamic auditory processing and phonological processing: Evidence from fMRI. *Journal of Cognitive Neuroscience, 13*(5), 687-697.

Poldrack, R. A., Wagner, A. D., Prull, M. W., Desmond, J. E., Glover, G. H., & Gabrieli, J. D. E. (1999). Functional specialization for semantic and phonological processing in the left inferior prefrontal cortex. *Neuroimage, 10*(1), 15-35.

Prabhakaran, R., Blumstein, S. E., Myers, E. B., & Hutchison, E. (submitted). An event-related fMRI investigation of phonological-lexical competition.

Price, C., Wise, R., Ramsay, S., Friston, K., Howard, D., Patterson, K., et al. (1992). Regional response differences within the human auditory cortex when listening to words. *Neuroscience Letters, 146*(2), 179-182.

Price, C. J. (1998). The functional anatomy of word comprehension and production. *Trends in Cognitive Sciences, 2*(8), 281-288.

Ravizza, S. M., Delgado, M. R., Chein, J. M., Becker, J. T., & Fiez, J. A. (2004). Functional dissociations within the inferior parietal cortex in verbal working memory. *Neuroimage, 22*(2), 562-573.

Repp, B. H. (1984). The role of release bursts in the perception of [s]-stop clusters. *J Acoust Soc Am, 75*(4), 1219-1230.

Rissman, J., Eliassen, J., & Blumstein, S. E. (2003). An event related fMRI investigation of implicit semantic priming. *Journal of Cognitive Neuroscience, 15*(8), 1160-1175.

Rorden, C., & Karnath, H. O. (2004). Using human brain lesions to infer function: a relic from a past era in the fMRI age? *Nat Rev Neurosci, 5*(10), 813-819.

Ryalls, J., Zipprer, A., & Baldauff, P. (1997). A preliminary investigation of the effects of gender and race on voice onset time. *Journal of Speech and Hearing Research, 40*(3), 642-645.

Saffran, E. M., Marin, O. S., & Yeni Komshian, G. H. (1976). An analysis of speech perception in word deafness. *Brain and Language, 3*(2), 209-228 URLJ.

Scott, S. K., Blank, C. C., Rosen, S., & Wise, R. J. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain, 123 Pt 12*, 2400-2406.

Scott, S. K., & Johnsrude, I. S. (2003). The neuroanatomical and functional organization of speech perception. *Trends Neurosci, 26*(2), 100-107.

Scott, S. K., & Wise, R. J. (2004). The functional neuroanatomy of prelexical processing in speech perception. *Cognition, 92*(1-2), 13-45.

Sharma, A., & Dorman, M. F. (1999). Cortical auditory evoked potential correlates of categorical perception of voice-onset time. *J Acoust Soc Am, 106*(2), 1078-1083.

Sharma, A., & Dorman, M. F. (2000). Neurophysiologic correlates of cross-language phonetic perception. *Journal of the Acoustical Society of America, 107*(5,Pt1), 2697-2703.

Sharma, A., Marsh, C. M., & Dorman, M. F. (2000). Relationship between N1 evoked potential morphology and the perception of voicing. *Journal of the Acoustical Society of America, 108*(6), 3030-3035.

Shivde, G., & Thompson-Schill, S. L. (2004). Dissociating semantic and phonological maintenance using fMRI. *Cogn Affect Behav Neurosci, 4*(1), 10-19.

Shulman, G. L., Corbetta, M., Buckner, R. L., Fiez, J. A., Miezin, F. M., Raichle, M. E., et al. (1997). Common blood flow changes across visual tasks: I.: Increases in subcortical structures and cerebellum but not in nonvisual cortex. *Journal of Cognitive Neuroscience, 9*(5), 624-647.

Spivey, M. J., Tanenhaus, M. K., Eberhard, K. M., & Sedivy, J. C. (2002). Eye movements and spoken language comprehension: effects of visual context on syntactic ambiguity resolution. *Cognit Psychol, 45*(4), 447-481.

Steinschneider, M., Schroeder, C. E., Arezzo, J. C., & Vaughan, H. G., Jr. (1994). Speech-evoked activity in primary auditory cortex: effects of voice onset time. *Electroencephalogr Clin Neurophysiol, 92*(1), 30-43.

Steinschneider, M., Schroeder, C. E., Arezzo, J. C., & Vaughan, H. G., Jr. (1995). Physiologic correlates of the voice onset time boundary in primary auditory cortex (A1) of the awake monkey: temporal response patterns. *Brain Lang, 48*(3), 326-340.

Stevens, K. N., & Blumstein, S. E. (1978). Invariant cues for place of articulation in stop consonants. *J Acoust Soc Am, 64*(5), 1358-1368.

Takahashi, N., Kawamura, M., Shinotou, H., Hirayama, K., Kaga, K., & Shindo, M. (1992). Pure word deafness due to left hemisphere damage. *Cortex, 28*(2), 295-303.

Talairach, J., & Tournoux, P. (1988). *A co-planar stereotaxic atlas of a human brain.* Stuttgart: Thieme.

Talavage, T. M., Edmister, W. B., Ledden, P. J., & Weisskoff, R. M. (1999). Quantitative assessment of auditory cortex responses induced by imager acoustic noise. *Human Brain Mapping, 7*(2), 79-88.

Tallal, P., & Piercy, M. (1975). Developmental aphasia: The perception of brief vowels and extended consonants. *Neuropsychologia, 13*, 69-74.

Tanaka, Y., Yamadori, A., & Mori, E. (1987). Pure word deafness following bilateral lesions. A psychophysical analysis. *Brain, 110*(Pt 2), 381-403.

Thompson-Schill, S. L., D'Esposito, M., Aguirre, G. K., & Farah, M. J. (1997). Role of left inferior prefrontal cortex in retrieval of semantic knowledge: a reevaluation. *Proc Natl Acad Sci U S A, 94*(26), 14792-14797.

Thompson-Schill, S. L., D'Esposito, M., & Kan, I. P. (1999). Effects of repetition and competition on activity in left prefrontal cortex during word generation. *Neuron, 23*(3), 513-522.

Utman, J. A., Blumstein, S. E., & Sullivan, K. (2001). Mapping from sound to meaning: Reduced lexical activation in Broca's aphasics. *Brain and Language, 79*(3), 444-472.

Van de Moortele, P. F., Cerf, B., Lobel, E., Paradis, A. L., Faurion, A., & Le Bihan, D. (1997). Latencies in fMRI time-series: effect of slice acquisition order and perception. *NMR Biomed, 10*(4-5), 230-236.

van Veen, V., Cohen, J. D., Botvinick, M. M., Stenger, V. A., & Carter, C. S. (2001). Anterior cingulate cortex, conflict monitoring, and levels of processing. *Neuroimage, 14*(6), 1302-1308.

Varney, N. R., Damasio, H., & Adler, S. (1989). The role of individual difference in determining the nature of comprehension defects in aphasia. *Cortex, 25*(1), 47-55.

Vitevitch, M. S., & Luce, P. A. (2004). A web-based interface to calculate phontactic probability for words and nonwords in English. *Behavior Research Methods, Instruments & Computers, 36*, 481-487.

Vitevitch, M. S., Luce, P. A., Pisoni, D. B., & Auer, E. T. (1999). Phonotactics, neighborhood activation, and lexical access for spoken words. *Brain Lang, 68*(1-2), 306-311.

Volaitis, L. E., & Miller, J. L. (1992). Phonetic prototypes: Influence of place of articulation and speaking rate on the internal structure of voicing categories. *Journal of the Acoustical Society of America, 92*(2, Pt 1), 723-735.

Wagner, A. D., Pare-Blagoev, E. J., Clark, J., & Poldrack, R. A. (2001). Recovering meaning: left prefrontal cortex guides controlled semantic retrieval. *Neuron, 31*(2), 329-338.

Wang, E., Peach, R. K., Xu, Y. S., Schneck, M., & Manry, C. (2000). Perception of dynamic acoustic patterns by an individual with unilateral verbal auditory agnosia. *Brain and Language, 73*, 442-255.

Warburton, E., Wise, R. J. S., Price, C. J., Weiller, C., Hadar, U., Ramsay, S., et al. (1996). Noun and verb retrieval by normal subjects studies with PET. *Brain, 119*, 159-179.

Warren, R. M. (1970). Perceptual restoration of missing speech sounds. *Science, 167*(3917), 392-393.

Wise, R. J., Greene, J., Buchel, C., & Scott, S. K. (1999). Brain regions involved in articulation. *Lancet, 353*(9158), 1057-1061.

Wise, R. J., Scott, S. K., Blank, S. C., Mummery, C. J., Murphy, K., & Warburton, E. A. (2001). Separate neural subsystems within 'Wernicke's area'. *Brain, 124*(Pt 1), 83-95.

Zarahn, E., Rakitin, B., Abela, D., Flynn, J., & Stern, Y. (2004). Positive Evidence against Human Hippocampal Involvement in Working Memory Maintenance of Familiar Stimuli. *Cereb Cortex.*

Zatorre, R. J., & Belin, P. (2001). Spectral and temporal processing in human auditory cortex. *Cerebral Cortex, 11*(10), 946-953.

Zatorre, R. J., Belin, P., & Penhune, V. B. (2002). Structure and function of auditory cortex: music and speech. *Trends in Cognitive Sciences*, *6*(1), 37-46.

Zatorre, R. J., Evans, A. C., Meyer, E., & Gjedde, A. (1992). Lateralization of phonetic and pitch discrimination in speech processing. *Science*, *256*(5058), 846-849.

Zatorre, R. J., Meyer, E., Gjedde, A., & Evans, A. C. (1996). PET studies of phonetic processes in speech perception: Review, replication, and re-analysis. *Cerebral Cortex*, *6*, 21-30.

Zurowski, B., Gostomzyk, J., Gron, G., Weller, R., Schirrmeister, H., Neumeier, B., et al. (2002). Dissociating a common working memory network from different neural substrates of phonological and spatial stimulus processing. *Neuroimage*, *15*(1), 45-57.

Appendix A. Distribution of stimuli into trial onset asynchrony bins, Experiment 3

	TOA bin						
Stimulus	2 sec	4 sec	6 sec	8 sec	10 sec	12 sec	**Total**
Gift1	4	3	3	4	3	3	**20**
Gift2	3	4	3	3	4	3	**20**
Gift3	3	3	4	3	3	4	**20**
Gift4	4	3	3	4	3	3	**20**
Gift5	3	4	3	3	4	3	**20**
Gift6	3	3	4	3	3	4	**20**
Gift7	3	4	3	3	4	3	**20**
Giss1	4	3	3	4	3	3	**20**
Giss2	3	4	3	3	4	3	**20**
Giss3	3	3	4	3	3	4	**20**
Giss4	4	3	3	4	3	3	**20**
Giss5	3	4	3	3	4	3	**20**
Giss6	3	3	4	3	3	4	**20**
Giss7	3	4	3	3	4	3	**20**
High	4	3	3	4	3	3	**20**
Low	3	3	4	3	3	4	**20**
Total	**53**	**54**	**53**	**53**	**54**	**53**	**320**