Adaptive paradigms for identifying the neural substrates of phonological encoding in individuals with and without aphasia

By

Melodie Yen

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Hearing and Speech Sciences

August 9, 2019

Nashville, Tennessee

Approved:

Stephen M. Wilson, Ph.D

Melissa Duff, Ph.D

Michael de Riesthal Ph.D., CCC-SLP

Sarah Brown-Schmidt, Ph.D

Acknowledgments

This work would not have been possible without the guidance of my advisor, Dr. Stephen Wilson. I would also like to thank Drs. Melissa Duff, Michael de Riesthal, Sarah Brown-Schmidt, and C. Melanie Schuele, who have given me invaluable feedback as part of my dissertation and advisory committees throughout my graduate career.

I am extremely grateful for the members of the Language Neuroscience Lab at Vanderbilt University Medical Center, who have made innumerable contributions to my work as my colleagues and labmates. I would also like to express my appreciation for the individuals who participated in my study.

This project was supported by the National Science Foundation Graduate Research Fellowship, the University of Arizona Fellows award, and grants R01 DC013270 and R21 DC016080 through the National Institutes of Health.

TABLE OF CONTENTS

Page

Chapter

LIST OF TABLES

Table *Page*

LIST OF FIGURES

Figure *Page*

CHAPTER I

INTRODUCTION

OVERVIEW AND PURPOSE

In most neurologically normal individuals, the core language network consists of frontal and temporal language regions, which are lateralized to the left hemisphere (Knecht et al., 2003; Seghier et al., 2011; Springer et al., 1999; Tzourio-Mazoyer et al., 2010). Additionally, the supramarginal gyrus and ventral precentral gyrus are specifically driven by phonological encoding (Gitelman et al., 2005; McDermott et al., 2003; Price et al., 1997), and more extensive temporal lobe regions are involved in semantic processing (Binder et al., 2009).

Identifying language processing regions at the level of individual participants is useful for at least three reasons. First, in the clinical context of presurgical language mapping, it would be valuable to be able to reliably activate these regions, especially in patients with surgical sites planned in the vicinity of these regions. The most effective existing language mapping paradigms are semantically based, and activate left posterior temporal cortex and a left frontal region that is anterior to the region involved in phonological encoding (Binder et al., 2008; Gaillard et al., 2007; Wilson et al., 2018). Commonly used tasks that might seem to entail phonological encoding, such as picture naming, do not actually activate these regions (Wilson et al., 2017, 2018). Second, identified areas could serve as regions of interest (ROIs) in studies that use “functional localizers” followed by probes of functional specialization (Fedorenko et al., 2010). Third, in the context of investigating functional reorganization of language processing in post-stroke aphasia, it would be beneficial to be able to identify language regions in individuals. For example, patients with aphasia often show a significant degree of recovery from phonological encoding deficits (Pashek & Holland, 1988; Laska et al., 2001; Yagata et al., 2017), suggesting that reorganization is possible, but tracing any patterns of reorganization depends on being able to reliably identify these regions across time.

Being able to determine with statistical rigor whether the localization of language regions changes over time is critical to the line of research concerned with functional reorganization (Kiran et al., 2013; Meinzer et al., 2013; Wilson et al., 2017). Language areas of the brain can be identified with functional magnetic resonance imaging (fMRI) using language mapping paradigms, which generally contrast conditions that involve language processing to conditions that do not (Binder et al., 2008). To support research on functional reorganization of language regions in recovery from aphasia, a language mapping paradigm that is capable of identifying language regions consistently across time and sensitively among participants is crucial.

FUNCTIONAL REORGANIZATION OF LANGUAGE

Recovery from aphasia after damage to language regions of the brain is thought to depend on neural plasticity, that is, functional reorganization of surviving brain regions such that they take on new or expanded roles in language processing (Crinion and Price, 2005; Geranmayeh et al., 2014; Heiss et al., 1997, 1999; Nadeau, 2014; Saur et al., 2006; Saur and Hartwigsen, 2012; Turkeltaub et al., 2011). There is currently a great deal of interest in characterizing the nature of this putative process of functional reorganization. A better understanding of when and how reorganization takes place, how different patterns of reorganization depend on patient-specific factors, and how different patterns are associated with better or worse outcomes, could inform the design of new therapies, and could facilitate optimal targeting of specific interventions to individual patients (Fridriksson et al., 2012; Thompson and den Ouden, 2008).

Heiss and Thiel (2006) proposed a hierarchical organization for such language recovery based on imaging and brain stimulation data. They observed from clinical and neuroimaging studies of fluent (Musso et al., 1999; Weiller et al., 1995) and non-fluent (Rosen et al., 2000) aphasic patients that activation of ipsilateral structures seemed to underlie the most favorable degree of recovery. Integrating this evidence with studies that use PET to measure changes in cerebral glucose metabolism during speech tasks (Heiss et al., 1993; Heiss et al., 1999; Warburton et al., 1999), they further perceived that right hemisphere regions were sometimes activated, but improvement nevertheless primarily depended on preserved components of the original language network. Studies using repetitive transcranial magnetic stimulation (rTMS) have attempted to elucidate the role of this right hemisphere activation by suppressing the homotopic pars triangularis in chronic non-fluent aphasia patients over a series of sessions (Martin et al., 2004; Naeser et al., 2005). These authors reported improved picture naming performance in this population after this procedure, suggesting that, consistent with a PET study that examined chronic patients undergoing melodic intonation therapy (Belin et al., 1996), right hemisphere activation may reflect inefficient or maladaptive attempts to support language function. Based on this pattern of data, Heiss and Thiel proposed the following hierarchy of recovery: first, the best level of recovery depends on restoration of original network function in the dominant hemisphere, likely only possible in cases of small lesions; second, activation of areas around the site of lesion supports some moderate but typically satisfactory degree of recovery, if primary language regions are damaged; finally, given a severely compromised dominant-hemisphere network, activation of contralateral homotopic areas contributes to language function, albeit less efficiently than mechanisms of ipsilateral compensation.

The role of the right hemisphere in recovery of language function remains contested. Studies such as Weiller et al. (1995) and Leff et al. (2002) using PET show that the right STG and posterior STS are differentially recruited by patients with aphasia performing active (pseudoword repetition, verb generation) and passive (monosyllabic speech listening) language tasks, respectively. Other studies have characterized this right hemisphere activity as reflecting compensatory function that in some cases seems essential to the patient's residual language ability, but nevertheless lacks the effectiveness of restoration of left hemisphere language areas (Winhuisen et al., 2005, 2007). Rosen et al. (2000) further illustrate this idea by showing that patients exhibit more activation than normal in the right IFG during attention-demanding lexical task, but the magnitude of this activation did not correlate with verbal performance. However, the perilesional response in the left IFG of the two best-performing patients did, suggesting that activity in spared or restored parts of the original language network is more functionally relevant than activity in contralateral homotopic areas, since it is associated with improved performance. This is consistent with Heiss and Thiel's depiction of recovery mechanisms as differentially effective, whereby language function is better recovered the more the brain is able to achieve its original functional organization.

In a more recent review, Hartwigsen and Saur (2017) again assessed the neuroimaging literature to date with respect to language recovery. The central themes of Heiss and Thiel's framework of recovery (hierarchical engagement of primary language regions, perilesional sites, and the contralateral hemisphere during recovery) recur in Hartwigsen and Saur's discussion, indicating that evidence for these concepts have persisted over time.

Hartwigsen and Saur additionally integrate new insights from neuroimaging and pharmacology to elaborate on our current understanding of functional recovery. For example, complementing the history of attempts to associate right hemisphere function with behavioral change, recent studies have investigated right hemisphere structure in relation to recovery of language production abilities. Xing et al. (2016) used voxel-based morphometry to show that right temporoparietal grey matter volume was greater in patients with aphasia compared to individuals with no history of aphasia, and this relative increase was positively correlated with scores on spontaneous speech, naming, and repetition tasks. Further evidence comes from Hope et al. (2017), who demonstrated two clusters in the right middle temporal gyrus in which increased grey matter volume correlated with improvement in picture naming. Interestingly, the 28 individuals included in this study were chronic patients with aphasia (first scanned at least one year post-stroke, mean 50.7 months) who were assessed twice (mean 30.7 months apart) to quantify any behavioral or neurological change. The acute and subacute phases are typically more inclined to show change; however, these authors demonstrated that structural changes may still be taking place in the chronic phase. While behavioral profiles were fundamentally stable at the group level, the changes observed in picture naming at the individual level were systematically related to changes in other word-finding tasks (written object naming and action naming) and with the changes in grey matter volume. This suggests that these differences are not spurious fluctuations in measurement noise, and as such, some degree of real spontaneous change may still occur chronically. Findings from both these studies persisted despite controlling for age, lesion size, time post-stroke, and baseline performance, indicating that structural properties of the right hemisphere remain relevant to recovery and its function may not be entirely maladaptive. Other more recent neuroimaging studies have espoused the importance of white matter tracts in explaining functional loss (Corbetta et al., 2015; Kümmerer et al., 2013) and predicting language outcomes (Kim et al., 2013; Forkel et al., 2014). Perhaps most excitingly, current technology makes it feasible to collect and analyze data accrued over the years. Projects such as the PLORAS database (Seghier et al., 2016), which includes behavioral and imaging data from over 760 stroke patients, will allow researchers to identify patterns that are consistent over time and potentially predict recovery outcomes from single scans or assessments.

THE ROLE OF EFFORT IN LANGUAGE PROCESSING

Task difficulty effects

Geranmayeh, Brownsett, and Wise (2014) discussed the role of domain-general systems in functional recovery after brain damage. In a review of the literature regarding patients with aphasia, they argued that much of the neuroimaging data taken as evidence of language reorganization is confounded by task difficulty effects. One study by Saur et al. (2006), for instance, exemplifies the nature of the problem. These authors collected fMRI data from 14 patients with aphasia while they performed language tasks at three different stages of recovery: acute (within days post-stroke), subacute (2 weeks), and chronic (approximately 1 year). One task was easy, comprised of a simple distinction between sentences played normally and sentences in reverse. The other task—detection of sentences with semantic violations—was more difficult. The behavioral results showed that, in the acute stage, few patients were able to detect semantic violations and focused their efforts on the simpler task instead. However, at 2 weeks post-stroke, the error rates suggested that almost all patients were attempting the more difficult task, despite their inability to reach perfect accuracy. Finally, at 1 year post-stroke, patients and healthy participants were similar in their performance on the semantic violation task. Interpretation of these results was critically driven by the activity observed in the right insular/inferior frontal gyrus over this same time course. Activation in this area was low in the acute phase, greater than normal at 2 weeks, then as low as healthy participants' by 1 year. Saur et al. interpreted this pattern as reflecting a language-specific process contributing to recovery, in which right hemisphere homotopic areas are recruited when left-hemisphere language regions are damaged. However, Geranmayeh and colleagues observed that several aspects of the study point to an alternative explanation. First, in the acute phase, a composite out-of-scanner language score was correlated with activity in both the left and right IFG. Second, improvement in this language score was correlated with not only the right anterior insular/IFG, but also the dorsal anterior cingulate cortex/superior frontal gyrus. This is an important component of the cingulo-opercular network, a system responsible for goal-directed functions, maintaining cognitive faculties relevant to current task demands, and salience detection (Fig. 1.1, red). These findings indicate that, while language-specific changes cannot be conclusively discounted, it is equally if not more plausible that improved performance is being supported by upregulation and subsequent normalization of domain-general networks. The trajectory of change, along with the full constellation of activated areas, suggests that engagement of the cingulo-opercular network reflects the effort required from the patients to perform the difficult task at each stage of recovery: low effort when the task is too challenging to consider, increased effort when partial recovery makes it plausible to attempt, then reduced effort when enough recovery had taken place to make the task simpler.

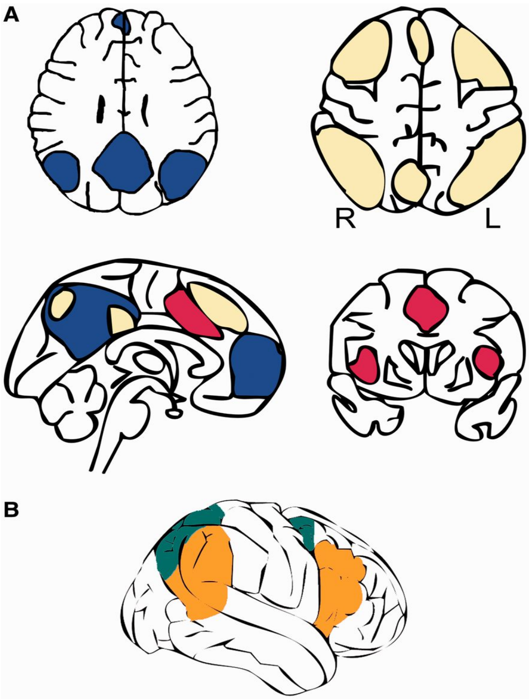


Figure 3 Schematic drawing of the typical spatial distribution of domain-general networks that may be engaged during neuroimaging of language tasks in healthy controls as well as aphasic patients. Many functional neuroimaging studies depict these networks as spatially overlapping. (**A**) The coloured networks are the Default Mode Network in blue, the fronto-parietal control network in yellow, and the cingulo-opercular network in red. The Default Mode Network is a 'task-negative' network that is deactivated during task performance on stimuli. Although they are functionally separable networks, the fronto-parietal control and cingulo-opercular networks often co-activate (see Fig. 2), and are considered to exert attention and executive control, and other processes involved in making a decision, selecting a response, and monitoring and correcting for errors. (**B**) Attentional networks can be divided into two broad systems; the dorsal attention network, in green, is thought to be a goal-driven 'top–down' attentional system, and is distributed symmetrically between the two hemispheres. The ventral attention network, in orange, is considered a stimulus-driven or 'bottom–up' attentional system, and largely lateralized to the right hemisphere.

Figure 1.1 Reproduced from Geranmayeh et al. (2014): Domain-general networks implicated in language processing.

As Geranmayeh and colleagues further discuss, functional reorganization is not the only domain in which interpretation is potentially confounded by effort effects. Our understanding of the typical function of Broca's area, for example, has evolved with a growing body of neuroimaging and patient research. Many studies have traditionally implicated this region, commonly defined as the pars triangularis and opercularis, in syntactic processing (Stromswold et al., 1996; Caramazza & Zurif, 1976). In conjunction with the observation that patients with damage to this region often present with grammatical deficits (Broca, 1861; Dax, 1863), Broca's area has been eminently perceived as a language region responsible for sentence structure. However, data demonstrate that damage confined to left inferior cortex does not result in lasting agrammatism (Mohr, 1976; Baldo and Dronkers, 2006), and moreover, Broca's area does not respond exclusively to linguistic stimuli. Imaging studies have frequently pointed to a role for the IFG in working memory (Smith et al., 1998; Fiez et al., 1999; Jonides et al., 1998; Braver et al.; 1997). Koechlin and Jubault (2006) showed that both left and right IFG were activated for tasks requiring rule-governed and hierarchically sequenced behaviors. Fedorenko et al. (2012) further demonstrated that subregions of Broca's area have different response profiles. While a language-selective subcomponent was found spanning the BA44/45 boundary, it was abutted anteriorly and posteriorly by a second subregion that responded to all tasks regardless of type (verbal, non-verbal, arithmetic, spatial working memory, cognitive control). Activity in this region was modulated not by the linguistic vs. non-linguistic distinction, but by task difficulty—it responded more to harder than easier conditions across all categories of stimuli. These results are consistent with a characterization of Broca's area as having a domain-general role in hierarchical organization modulated by difficulty, potentially with a smaller embedded subsystem optimized for the sequencing required in language processing. This finding that the IFG is sensitive to task difficulty illustrates the confound that arises when effort is involved in the performance of a task intended to elicit language-related activity. Activation produced by the task is sometimes ascribed to the domain-specific (e.g. linguistic) component of the task, when it should be attributed to the domain-general process of exerting effort to perform the task, or to an interaction between the two systems.

Resolving the task difficulty confound

Controlling effort between experimental and control conditions

Geranmayeh and colleagues make a number of recommendations for addressing the issue of task difficulty effects in the context of language research. First is the comparison of the linguistic condition with an equally demanding non-linguistic control condition. This is necessary to ensure that any difference in the activation produced by the two conditions is indeed due to the manipulation of interest, rather than to the fact that one condition requires more effort than the other to perform.

The passive rest condition has often been used as a baseline control, but presents with both benefits and drawbacks. Rest requires no response and poses minimal demands on the participant, affording it the advantage of feasibility for use with language-impaired populations and as a low-level baseline for sensory studies. However, its usefulness in a subtractive design may be limited given the upregulation of the default mode network during rest (Spitsyna et al., 2006; Awad et al., 2007). Binder et al. (1999) discuss this issue in a neuroimaging study designed to characterize the activity and organization of this network. They found that the resting state network involved many of the same neural substrates underlying the semantic system, including the left angular gyrus, posterior cingulate, dorsal prefrontal cortex, and parahippocampus when compared to a tone task designed to interfere with passive task-independent thought processes. The authors interpreted the sharing of these areas as supporting internal processes that include semantic content (such as memories, plans, maintaining an internal monologue, and long-term problem solving). Binder et al. (2009) reinforced the idea that human brains are not simply inactive during “rest” in a meta-analysis of 120 functional imaging studies. This review consolidated a vast body of fMRI and PET research to localize the human semantic system to a network of regions including the angular gyrus, inferior frontal gyrus, lateral and ventral temporal cortex, dorsal medial prefrontal cortex, ventromedial prefrontal cortex, and posterior cingulate gyrus. Again, they emphasized the architectural similarities between the semantic and resting state network, and reiterated the observation that task-unrelated thoughts are fundamentally semantic in nature. As such, using passive rest as a baseline comes with the caveat of subtracting away semantic-related activation when contrasted with another condition. Binder et al. (1999) observed that this activation produced at rest was generally bilateral but showed some left hemisphere dominance. This is consistent with studies showing left-lateralization of processes such as decision-making (Vallar et al., 1988), compositionally generating mental images (Farah et al., 1985; Kosslyn, 1988), and making rapid inferences (Gazzaniga, 1995). The lateralization of internally-generated activity suggests that these cognitive processes may be mediated by language. In sum, activity during rest is not random and non-localized, but rather organized and systematically regulated as a function of the nature of internally-driven processes. This natural fluctuation in the default mode network makes passive rest a sub-optimal baseline condition in some situations—for instance, tasks involving problem-solving may have their activation disguised by the comparison to internally-directed thoughts during rest. Suppression of the network as a result of task engagement may also be misinterpreted as meaningful deactivation induced by the condition of interest. Given that passive rest will engage a network inherent to task-independent processes, using a higher-level baseline task is often useful. The control condition would then match the condition of interest on as many variables as possible, ideally differing only in the critical variable of manipulation.

Controlling effort between participants

Second, when comparing language activation between two groups, it is necessary to ensure that both groups found the tasks equally challenging. Just as no condition should be more difficult than another within a task paradigm, no paradigm should be more difficult for one subset of participants than another. This idea was previously espoused by Price, Crinion, and Friston (2006) in a discussion of how neuroimaging studies of neurologically impaired individuals should be approached. The authors put forth several recommendations for imaging in this population, including minimizing inter-stimulus intervals as appropriate and blocking stimulus types to capitalize on the summation of hemodynamic response peaks. Like Geranmayeh et al. (2014), they stressed the primary concern of employing tasks that patients can perform. This is a particularly challenging criterion to meet with respect to research involving individuals with language impairments. Given the nature of their deficits, patients with aphasia may find even a simple language task difficult or impossible to perform, whereas a control group of neurotypical participants would find the same task trivially easy. This disproportionate amount of effort required from the two groups would therefore leave the results vulnerable to misinterpretation—for example, difficulty-related activity increase in the IFG (c.f. Fedorenko et al., 2012) for the patient group that is absent in the healthy group may be erroneously interpreted as functional compensation or reorganization.

Addressing this issue has critical implications for study design. One approach is to use a task in which patients can perform at accuracy levels matching those of controls. Paradigms such as passive speech comprehension are frequently used because they do not require the participant to respond, and as such are feasible for many patient populations. This task typically consists of a language condition compared to an acoustically-matched control condition, such as backward speech. It is moderately successful in activating language areas in a reliable way—however, this activation tends to be limited to temporal regions and shows little sensitivity to frontal components of language processing (Harrington et al., 2006; Crinion et al., 2003; Wilson et al., 2017). Picture naming is frequently used in language research as well, given its simple implementation and response demands. In this task, participants are presented with pictures of objects and asked to name them aloud. Visually-matched stimuli (such as scrambled versions of the same pictures) are typically used as control. This paradigm can be useful in that trials with correct and incorrect responses can be analyzed separately (Fridriksson et al., 2009). While picture naming shows moderately strong reliability, this advantage of the paradigm is largely due to the bilateral sensorimotor activation associated with the production component of the task, and is relatively uninformative in terms of language localization (Harrington et al., 2006; Rau et al., 2007; Wilson et al., 2017). Psychometric shortcomings aside, such tasks that are simple enough for patients to match neurotypical performance will nevertheless exclude many moderately to severely impaired patients, compromising the applicability of the paradigm and the generalizability of the results. Moreover, a simple task would place relatively low cognitive demands on healthy control participants. Even if both groups achieve the same level of accuracy on a task, the amount of effort required to do so may be less for controls compared to patients.

An alternative approach is to use a task with a more difficult version for controls. However, this option presents with limitations as well. Speech comprehension paradigms lend themselves to more or less challenging versions fairly readily, but not so necessarily for paradigms with an active component. Raboyeau et al. (2008) compared object naming in aphasia patients to object naming in a foreign language for controls after a training period. While this design successfully makes the naming task more challenging for healthy participants, it also invites confounds from comparing processes involved in two similar but non-identical tasks (in this case, impaired L1 naming compared to partially learned L2 naming). Gaillard and colleagues (2004, 2007) developed an Auditory Definition Decision Task in which participants hear blocks of forward sentences, backward sentences, and silence. They pressed a button if the sentence was true (forward condition) or if they heard a beep (backward condition). The task approximated difficulty between groups and individuals by proceeding on one of several difficulty levels that determined the lexical frequency of the target word (e.g. *chair* in *Something you sit on is a* *chair*). This paradigm has shown some success in localizing language regions in healthy controls (Gaillard et al., 2007; Berl et al., 2014) and in patients of perinatal stroke (Newport et al., 2017). The ADDT shows advantages over many other existing paradigms (such as the narrative comprehension and picture naming tasks described previously) in that it is left-lateralizing and fairly sensitive to the core frontal and temporal language regions. Moreover, the three or four difficulty levels (depending on task version) attempt to hold constant any task difficulty effects. However, difficulty is manipulated through the single dimension of lexical frequency of the target word in the forward speech condition, and not at all in the backward speech condition. In their implementation of the ADDT, Berl et al. (2014) pre-determined which difficulty levels would be developed for which subsets of their participants, based on age group rather than more personalized factors such as individual ability.

Geranmayeh et al. emphasize the importance of assessing error rates and reaction times (RTs) to measure the extent to which difficulty is matched across participant groups. Given the limited success of the field in achieving this matching, they make some recommendations for alternatives: including the corresponding variables in the regression model can control for these effects in analysis; additionally, discarding errorful volumes is a viable option for event-related fMRI. Some studies, such as Binder et al. (2005), incorporate difficulty-induced activity into their results. This study was primarily concerned with investigating lexical mechanisms (by contrasting word regularity, lexicality, and concreteness), but included analyses of the results with and without taking RTs into account in the regression model. When the effect of RTs across stimulus conditions was examined, they found that time on task was positively correlated with activity in the cingulo-opercular and fronto-parietal networks. This pattern was consistent with a previous study conducted by some of the same authors, designed to localize the neural correlates of auditory object identification speed and accuracy. In response to speech sounds masked with noise to varying degrees, fMRI activation showed that inferior frontal, anterior insular, and frontal opercular cortices predicted RT (Binder et al., 2004). Wilson et al. (2016) also capitalized on the construct of reaction time by using it as a proxy for syntactic processing load on each trial. They presented 51 primary progressive aphasia patients and 24 healthy controls with seven sentence conditions varying in levels of syntactic complexity. The participants performed a sentence-picture matching task with these stimuli while being scanned with fMRI. Results showed that atrophy of the left inferior frontal cortex (previously established to be associated with the syntactic deficits observed in PPA) was predictive of signal modulation as a function of syntactic demands in a more widespread fronto-temporo-parietal network. These studies illustrate the usefulness of collecting behavioral measures such as RTs—they can not only serve as an indicator of success in matching task difficulty across participants, but also be analyzed as markers of particular cognitive processes.

To further investigate and interpret the role of reaction time measures in language research contexts, Taylor, Rastle, and Davis (2014) had participants read lists of real words, irregular words, and pseudowords out loud during fMRI. They found that certain areas responded more to words than pseudowords (including the left angular and middle temporal gyri) and others responded more to pseudowords than words (including bilateral IFG, precentral gyrus, superior and inferior parietal cortices, supplementary motor cortex and left inferior occipital cortex) even when RT effects were accounted for. On the other hand, a widespread network of regions was correlated with RT for all stimulus item types, including the bilateral occipitotemporal and parietal cortices, inferior frontal and precentral gyri, insula, and middle temporal gyrus. By conducting several analyses examining the effects of including or excluding

RT as a covariate, the authors determined that RT should be independent of differences in neural activity between conditions if the corresponding brain regions are predicted to be sensitive to stimulus type (e.g. real words vs. pseudowords). However, RT *should* covary with differences in neural activity between conditions if the areas of interest represent both stimulus types, but one incurs less processing difficulty than the other. Taylor and colleagues describe this as a difference in neural engagement and neural effort. This perspective on measures of behavior aligns with their general view of principles of activation (Taylor et al., 2013): a brain region should be activated more by a stimulus that it represents than by one it doesn't, and of the stimuli that it does represent, one that well-fits these representations should produce less activation than one that fits less well. This difference is attributable to level of effort (ill-fitting stimuli place greater processing demands on the region responsible for its representation) and reflected in RTs (longer times covarying this with activity). Their framework constitutes a sophisticated perspective on the recommendations regarding RT posed by Geranmayeh et al. in that it views RT as neither solely a confound nor necessarily an index of the cognitive processes under study, but rather a behavioral measure whose variation should pattern systematically depending on theoretically-motivated predictions. These studies confirm the influence of task difficulty on language-related activation, highlighting the intricate intermingling of domain-general and language-specific systems in the brain, and demonstrating the importance of conscientious task design.

Current approach

Adaptive staircases have been used in the psychological sciences to determine perceptual and behavioral thresholds in individuals. This psychophysical procedure adjusts the stimulus intensity on a trial-by-trial basis according to the participant's performance. For example, if an individual succeeds in detecting a grating pattern on a screen during one trial, the contrast of the stimulus against the background is decreased on the next until the participant can no longer do so. Then, the contrast is increased until it is detectable again, and so on. With enough trials, this dynamic adjustment of stimulus steps converges on the participant's detection threshold for that particular stimulus. This procedure can also be applied to the context of task difficulty. We used an adaptive staircase design to construct a series of paradigms capable of activating language areas while holding task difficulty relatively constant. In our adaptive rhyme task, two pseudowords are presented on the screen and the participant judges whether or not they rhyme. Stimuli again spanned seven levels of difficulty but varied in phonological complexity through word length, stress pattern, and degree of similarity between the two pseudowords, in addition to presentation rate. In our adaptive semantic task, two words are presented on the screen and the participant judges whether or not they are semantically related. The stimuli ranged in difficulty along the dimensions of lexical frequency, concreteness, degree of relatedness, word length, age of acquisition, and presentation rate. The control condition for both of these tasks consists of a perceptual matching task, in which participants judge whether or not two false font strings are identical. All of these tasks are structured as 2-down-1-up adaptive staircases with weighted step sizes. This design, whereby two consecutive correct responses are required to increase the difficulty of the next trial by one level, and one incorrect response decreases the difficulty by two levels, should converge at approximately 80% accuracy (García-Pérez, 1998).

These paradigms are highly promising with respect to their potential for application in language research. They address Geranmayeh et al.'s recommendations in the following ways:

(1) the adaptive design tailors the difficulty to each participant, making the task demanding but achievable for individuals of varying degrees of language competence (controlling for effort between participant groups)

(2) both the language and perceptual control conditions are adaptive (controlling for effort between conditions within a paradigm)

(3) accuracy, difficulty level, and reaction time measures are collected for analysis.

CHAPTER II

Experiment I: Paradigm development

INTRODUCTION

Existing language mapping paradigms fail to consistently localize brain regions involved in language processing in individual participants. The goal of the project is to develop tasks capable of doing so, for the purposes of guiding surgical resection of cortex, improving functional localizer precision, and informing language recovery research. Identifying phonological processing regions at the individual level would, in particular, constitute progress toward these ends.

Phonological encoding is an important stage of speech production in which lexical forms are mapped onto articulatory plans, involving processes such as selection and sequencing of phonemes, syllabification, and various phonological and morphophonemic operations (Levelt, 1989, 2001). Functional neuroimaging studies using tasks that entail phonological encoding have consistently identified two regions that are important for this process: the left supramarginal gyrus (SMG) and left posterior inferior frontal cortex (Paulesu et al., 1993, McDermott et al., 2003; Oberhuber et al., 2016; Price et al., 1997; Seghier et al., 2004; Twomey et al., 2015). The frontal activations have been reported in the posterior inferior frontal gyrus and the ventral precentral gyrus, but the latter in particular appears to be involved in phonological encoding specifically rather than language processing in general (Gitelman et al., 2005; McDermott et al., 2003; Price et al., 1997). Neuropsychological studies have shown that lesions to inferior parietal and/or posterior inferior frontal left hemisphere regions are associated with phonological encoding deficits (Baldo & Dronkers, 2006; Buchsbaum et al., 2011; Fridriksson et al., 2010, 2016; Geva et al., 2011; Mirman et al., 2015; Pillay et al., 2014). Further evidence for the role of these regions in phonological encoding comes from transcranial magnetic stimulation, which disrupts performance of phonological tasks when applied over the posterior inferior frontal gyrus (Gough et al., 2005; Hartwigsen et al., 2010; Grabski et al., 2013) or the supramarginal gyrus (Sliwinska et al., 2012; Hartwigsen et al., 2016).

While the neuroanatomical correlates of phonological encoding are therefore quite well established, it would be useful to be able to identify these phonological regions in individual participants, for at least three reasons. First, in the clinical context of presurgical language mapping, identification of these regions is critical in patients with surgical sites likely to be in their vicinity. The most effective existing language mapping paradigms are semantically based, and activate the left inferior frontal gyrus and left posterior temporal cortex (Binder et al., 2008; Gaillard et al., 2007; Wilson et al., 2018). The left-lateralized fronto-parietal phonological encoding regions are not identified by semantic paradigms, which is a significant limitation given the importance of these regions for speech production. Commonly used tasks that would seem to clearly entail phonological encoding, such as picture naming, also do not actually reliably activate these regions (Wilson et al., 2017, 2018). Second, it would be particularly useful to be able to identify phonological regions in the context of investigating functional reorganization of language processing in recovery from aphasia. Individuals with aphasia often show a significant degree of recovery from phonological encoding deficits (Pashek & Holland, 1988; Laska et al., 2001; Yagata et al., 2017), suggesting that reorganization is possible, but tracking any reorganization depends on being able to reliably identify these regions across time.

Third, phonological regions could serve as regions of interest (ROIs) in studies that use "functional localizers" followed by probes of functional specialization (Fedorenko et al., 2010).

We have previously shown that an adaptive semantic matching paradigm identifies frontal and temporal language areas in a reliable and valid manner in individual participants with and without language deficits (Wilson et al., 2018). The semantic paradigm involves making judgments on pairs of words as to whether they are semantically related. This task is compared to a perceptual control task that requires judgments on pairs of symbol strings. Critically, the difficulty of both tasks is adjusted on a trial by trial basis so that each participant finds the tasks challenging, but still within their competence. The effectiveness of the semantic paradigm appears to arise from the deep semantic processing it entails, as well as its adaptive nature.

The overall goal of the present study was to develop a paradigm for identifying phonological encoding regions based on the same principles. We designed, implemented and compared two different approaches for engaging the inferior parietal and inferior frontal phonological regions. The first was a rhyme judgment paradigm (Cousin et al., 2007; McDermott et al., 2003; Pugh et al., 1996; Seghier et al., 2004) and the second was a syllable counting paradigm (DeMarco, 2016; Poldrack et al., 1999; Price et al., 1997). Both tasks involved pseudowords in order to maximize dependence on phonological processing. Both tasks were intended as proxies for the phonological encoding stage of speech production. Neither involved overt speech production, for which it is very difficult to design effective control conditions (Braun et al., 1997). However, successful performance of both tasks required phonological encoding and could not be achieved based on orthographic strings alone.

We compared the two adaptive phonological paradigms to the previously described semantic paradigm in 16 neurologically normal individuals who were scanned on all three paradigms in one session. Our primary aim was to assess the validity and reliability of the two phonological paradigms. Validity was defined as the extent to which each paradigm activated the left-lateralized supramarginal and ventral precentral regions that are the established neural substrates of phonological encoding at the population level. The phonological paradigms were compared to each other, and to the semantic paradigm, which was not expected to activate these regions. Reliability was assessed by splitting each run in half and quantifying the overlap between activation maps derived from the two halves of each run.

METHODS

Participants

Sixteen neurologically normal adults participated in the experiment (age 57.0 ± 15.0 years, range 23–77 years); 6 male, 10 female; 12 right-handed, 3 left-handed, 1 ambidextrous; education 16.7 ± 2.2 years (range 12–20 years)). Participants were recruited mostly through a neighborhood listserv in Nashville, Tennessee. The inclusion criteria were: (1) fluent English speaker; (2) no history of neurological disease or speech/language or learning disabilities; (3) no contraindications for MRI. Patients were recruited in the approximate age range of a typical stroke population, given our intention of applying these procedures to individuals with aphasia in the future. Exclusion criteria were major psychiatric disorders, serious substance abuse, claustrophobia, or contraindications to MRI. Participants were administered the Mini-Mental State Examination (Folstein et al., 1975) and the Quick Aphasia Battery (Wilson et al., 2018) to confirm that their cognitive and language capacities were within normal limits (MMSE range: 27–30; QAB range: 9.31–10).

An additional twelve neurologically normal adults participated in a behavioral experiment for norming of stimuli (age 32.3 ± 8.6 years, range 25–51 years; 3 male, 9 female; 11 right-handed, 1 left-handed; education 17.3 ± 1.2 years, range 16–20 years). They were recruited similarly with the same inclusion and exclusion criteria, except that we were not concerned with contraindications for MRI, and we made no attempt to match the age range of a stroke population.

All participants gave written informed consent and were compensated for their time. The study was approved by the institutional review board at Vanderbilt University, and all study procedures were performed in accordance with the Declaration of Helsinki.

Adaptive rhyme judgment paradigm

The adaptive rhyme judgment paradigm comprised two conditions: a rhyme judgment task, and a perceptual matching task. The tasks were presented in alternating 20-s blocks in a simple AB block design. There were 10 blocks per condition, for a total scan time of 400 s (6:40). Each block contained between 4 and 10 items (inter-trial interval 5–2 s), depending on the level of difficulty (see below).

In the rhyme judgment condition, each item consisted of a pair of pseudowords, which were presented one above the other in the center of the screen. Half of the pairs rhymed (e.g. *mulky-tulkie*), while the other half did not (e.g. *shofy-sheffy*). Pairs were considered rhymes if their endings were phonologically identical starting from the rime of the syllable carrying primary stress. The participant pressed a button with a finger of their left hand if they decided that the pseudowords rhymed. If the words did not rhyme, they did nothing.

This task relies on phonological encoding because it requires the participant to assemble speech sounds into novel sequences, syllabify them, and determine a stress pattern, implicitly applying (morpho)phonological rules and conforming to phonotactic constraints. Pseudowords were used in order to load on phonological processes as exclusively as possible, minimizing engagement of lexical and semantic processes. It is important to note that rhyme judgments cannot generally be performed directly on orthographic representations, because rhyming depends on phonemic identity and stress placement, which are available only after phonological encoding. However it must be acknowledged that the rhyme task is not a perfect model of phonological encoding, because whereas the process of phonological encoding takes as its starting point something like a sequence of phonemes comprising the lexical word form, here the starting point is an orthographic string. This entails that there is also an orthographic decoding component to the task, which may be related to phonological encoding, but is presumably not equivalent to it. There is also the rhyme judgment itself, which is a metalinguistic operation on the output of phonological encoding.

The use of just a single button obviated the need for participants to learn an arbitrary association between ‘match’ and one button, and ‘mismatch’ and another button, which can be challenging in patient populations. The left hand was used for the button press to allow the paradigm to be applied in the future to individuals with post-stroke aphasia, many of whom have right-sided hemiparesis.

In the perceptual matching condition, each item comprised a pair of false font strings, presented one above the other. Half of the pairs were identical (e.g. ΔΘδѢЧ-ΔΘδѢЧ), while the other half were not identical (e.g. ΔΘδѢЧ-ЧΔɰƘΔ). The participant pressed the button if the strings were identical, and did nothing if they differed. The rhyme judgment and perceptual tasks were equivalent in terms of sensorimotor, executive and decision-making components, yet made differential demands on language processing. Task-switching demands were minimized because both conditions involved essentially similar tasks: pressing a button to matching pairs.

Critically, both the rhyme judgment task and the perceptual task were independently adaptive to participant performance. Each task had seven levels of difficulty. Whenever the participant made two successive correct responses on a given condition, they moved to the next highest level of difficulty on the subsequent trial of that condition. Whenever they made an incorrect response, they moved two levels down on the next trial. This is a 2-down-1-up adaptive staircase with weighted step sizes (up twice as large as down), which theoretically should converge at just over 80% accuracy (García-Pérez, 1998). Note that the difficulty level was manipulated independently for the rhyme and perceptual conditions, even though sets of items from the two conditions were interleaved due to the AB block design.

Manipulation of item difficulty

The difficulty of rhyme judgment items was manipulated by varying pseudoword length, orthographic transparency, stress pattern, and presentation rate (Table 2.1). On the first level of difficulty, all words were one syllable long (offering only a single option where stress can fall), and had very basic syllable shapes (CVC, CVV, or CVVC) and unambiguous orthography. For this easiest level, the rhyming portions of each matching pair were always orthographically identical (e.g. *zon-lon*, *heef-meef*). Most mismatching pairs shared onsets, with the mismatch created in the nucleus (*bool-beel*) or coda (*wod-wob*). Unlike all subsequent levels, rhyming pairs were more orthographically similar than mismatching pairs, which was a consequence of the restrictions described. We did not consider this a serious limitation, since we hoped to create a level on which even patients with phonological deficits might be capable of performing above chance, so a degree of orthographic transparency supporting performance was acceptable.

On the second level, all pseudowords were still one syllable long, but were composed of less straightforward shapes and orthography. For example, this level included pairs such as *doke-goak*, which are clear rhymes but depend on the knowledge that some sounds can be represented in multiple ways in English orthography. As such, unlike in the first level, rhyming pseudowords in the second level did not always have orthographically identical endings. This helped prevent participants from performing the task using only a superficial scan of letter similarity. To this same end, non-rhyming pairs often had letters in common (*plam-pran*). Every difficulty level from this level onward had approximately the same average Levenshtein distance, or the amount of letter mismatch between two strings, for rhyming pairs compared to non-rhyming pairs. This minimized the participants' ability to identify a pair as a match or mismatch by relying only on the resemblance of the word endings.

On the third level, all pseudowords were two syllables long and carried trochaic stress. As with the previous levels, mismatching pairs often shared letters to give the impression that they may potentially rhyme (*tarva-gava, doory-poroy*) and therefore ensure that the participant remained engaged. With greater allotments of letters and syllables, these later levels allowed matches to appear like mismatches (*boary-forrie*) and mismatches to appear like matches (*tiner-minner*) more readily. On the fourth level, pseudoword pairs consisted of two or mixed syllables, meaning the words could both be two syllables long (*sawby-tauby*), or one-syllable and two-syllables long (*kide-beride*). This level also incorporated some iambic stress, a less canonical stress pattern in English, to introduce further complexity into the items. On the fifth level, word pairs were composed of three (*paduka-barooka*) or mixed (*gimbo-pakimbo*) syllables. Primary stress fell on the penultimate syllable. With longer words, it is less immediately apparent where the stress might fall. By building primary stress into more middle segments of the words, the paradigm compels the participant to process the entire word, identify the most critical syllables out of several, then retain them for comparison with the next word. On the sixth level, word pairs consisted of three, four (*magorable-weplorable*), or mixed (*pendify-immendify*) syllables. Primary stress fell mostly on the antepenultimate syllable; some items carried penultimate stress. On the seventh and highest level, word pairs consisted of five (*inderbesticle-chiropestical*) or mixed (*jeechable-onumbeachable*) syllables. Primary stress fell on the antepenultimate syllable. In this last level, and to some extent the sixth level, many pseudowords incorporated morphemes (such as the endings -*ity* or -*ater*) to meet the greater length goal for the more difficult levels.

Table 2.1 Rhyming judgment levels of difficulty

Level	Length	Levenshtein distance (match)	Levenshtein distance (mismatch)	Match example	Mismatch example	Perceptual match	Perceptual mismatch
1	3.79 ± 0.41	1.00 ± 0.32	1.57 ± 0.88	*foo* *voo*	*soom* *soob*	ƽƱƽδΔ ƽƱƽδΔ	ƱΨЖδШΦ δŒΦƘƧ
2	3.77 ± 0.43	2.00 ± 0.95	1.63 ± 0.81	*scaw* *praw*	*pake* *pape*	ΘƟδΓΓ ΘƟδΓΓ	ƋΣƽδƟ ΘƧΘƟƧ
3	5.40 ± 0.97	3.00 ± 1.00	2.16 ± 0.75	*veery* *feary*	*shofy* *sheffy*	ƋΔδΦƱ ƋΔδΦƱ	ЧΓƽЧΨŒ ƧΓƽЧŒƽ
4	5.50 ± 0.78	3.25 ± 1.29	2.75 ± 0.97	*lameed* *abede*	*renerve* *setarve*	ŒΘƽƽƕ ŒΘƽƽƕ	δΘЧΦΨ ƧδЧΦΨ
5	7.03 ± 1.04	3.40 ± 1.14	3.05 ± 1.03	*loschito* *fuspeeto*	*lodacco* *topucko*	δδƱƱƘΦ δδƱƱƘΦ	δƧЧΞЧ ЧƧЧΞЧ
6	9.45 ± 1.07	4.60 ± 1.28	4.27 ± 1.08	*cupadimote* *recadimoat*	*vermissible* *vanisibe*	ШЧΞΞΦƽ ШЧΞΞΦƽ	ΨШΘЖƘΞ ΨƘΘŒƘΞ
7	11.58 ± 1.58	5.28 ± 1.31	4.92 ± 1.22	*cypermollicle* *waiperbollical*	*devapitulate* *rumipostulate*	ЖЧƽΞƽƕ ЖЧƽΞƽƕ	ΔƟЖƏŒ ΔƟΔƏŒ

Pseudowords were created by first consulting the UWA MRC psycholinguistic database (Coltheart, 1981; Wilson, 1988) for a list of real words at least three letters long and with a

Kucera-Francis frequency greater than one. These words were then input into the Wuggy pseudoword generator (Keuleers & Brysbaert, 2010) to derive approximately 25,000 pseudowords that matched the real-word inputs on the parameters of letter length, transition frequencies of graphemes, and retaining two thirds of the segments from the real words, making the results word-like but not readily recognizable as related to real words. Four hundred stimulus pairs were then created manually with some reference to this pseudoword list, but almost all final items underwent significant editing in order to fit the design goals for each difficulty level. This degree of manual editing for individual pseudoword pairs was necessary due to the specific circumstances that must be met for words to rhyme, and for non-rhyming words to look similar enough to pose a challenge.

Twelve neurologically normal adults were asked to make rhyme judgments on all 400 items, emphasizing accuracy over speed. Accuracy was 88.9% ± 5.5% (range 78.3–96.5%). Many of the incorrect responses seemed to represent genuine errors rather than ambiguities of how the pseudowords should be pronounced. For instance, 2 out of 12 participants judged *pake* and *pape* to rhyme, which is not possible since they differ in their codas. For 28 out of 400 items (7.0%), less than two thirds of responses were correct. These items were rechecked and 18 were deemed ambiguous or problematic. These items were included in the present imaging experiment, but will be excluded in future applications.

In the perceptual condition, item difficulty was manipulated in two ways: as the level of difficulty increased, mismatching pairs were more similar (Table 1), and presentation rate was faster. The length of the items (number of symbols) was dynamically matched to the length of the pseudoword pairs (number of letters) as the experiment progressed.

In order to match sensorimotor and executive demands across the language and perceptual conditions, it was necessary to yoke presentation rate across conditions. Presentation rate was adjusted at the start of each rhyme judgment block and remained fixed for the upcoming rhyme and perceptual blocks. The ‘ideal’ inter-trial interval for each condition was defined as the block length (20 s) divided by the ideal number of items per block (4 through 10, for difficulty levels 1 through 7). The number of items per block was then selected to be as large as possible without exceeding the average of the two ‘ideal’ inter-trial intervals (Wilson et al., 2018).

Adaptive syllable counting paradigm

The adaptive syllable counting paradigm was identical to the rhyme judgment paradigm is almost every respect. The only differences were in the pseudowords used and the nature of the decision to be made on them. In this paradigm, half of the pseudoword pairs had the same number of syllables (e.g. *lony-rado*), while the other half did not (e.g. *bosk-mipid*). The participant pressed a button if they decided that the pseudowords had the same number of syllables. If they did not have the same number of syllables, they did nothing.

This task entails phonological encoding because syllabification is not inherent in an orthographic string, but is implicitly computed based on phonological rules and phonotactic constraints. As for the rhyme task, there is also an orthographic decoding component to the task, as well as the counting and comparison of the number of syllables. Therefore like the rhyme task, the syllables task involves phonological encoding along with several other operations.

Manipulation of item difficulty

The difficulty of syllable counting items was manipulated by varying pseudoword length, the number of syllables in the pseudowords, and presentation rate (Table 2.2). The first level included pseudowords that were 3–4 letters long. Items in matching pairs were one syllable long, while mismatching pairs were one and two syllables long. The second level introduced more difficulty by allowing matching pairs to consist of either one-syllable or two-syllable pseudowords. Letter lengths remained the same and mismatching pairs continued to comprise one- and two-syllable pseudowords. The remaining third to seventh levels steadily increased in the number of letters and syllables involved, as shown in Table 2.

Stimuli were based on the same 25,000-pseudoword list generated for the rhyme paradigm. Items were excluded if they violated English phonotactics (e.g. *bathtr*), were pseudohomophones (e.g. *kee*), allowed for ambiguous pronunciation (e.g. *afed*), or appeared unusual in any other way (e.g. *awbix, rajue*). The final list of stimuli for this paradigm consisted of 4,273 pseudowords. Unlike in the rhyme judgment paradigm, it was not necessary to generate specific matching and mismatching pairs. Rather, matching and mismatching pairs were dynamically created from the 4,273-word list according to the current difficulty level.

Table 2.2 Syllable counting levels of difficulty

Level	Match	Mismatch	Match example	Mismatch example	Perceptual match	Perceptual mismatch
1	3 or 4 letters 1 syllable	3 or 4 letters 1 or 2 syllables	*fod* *bim*	*gan* *epo*	ƽƱƽδΔ ƽƱƽδΔ	ƱΨЖδШΦ δŒΦƘƧ
2	3 or 4 letters 1 or 2 syllables	3 or 4 letters 1 or 2 syllables	*gock* *tiss*	*sork* *boma*	ΘѲδΓΓ ΘѲδΓΓ	ƋΣƽδѲ ѲƧΘѲƧ
3	4 or 5 letters 1 or 2 syllables	4 or 5 letters 1 or 2 syllables	*lapo* *pany*	*namp* *ambus*	ƋΔδΦƱ ƋΔδΦƱ	ƔΓƽƔΨŒ ƧΓƽƔŒƽ
4	5 or 6 letters 1 or 2 syllables	5 or 6 letters 1 or 2 syllables	*tarmo* *vegar*	*blash* *ragle*	ŒΘƽƽƕ ŒΘƽƽƕ	δΘƔΦΨ ƧδƔΦΨ
5	6 or 7 letters 2 or 3 syllables	6 or 7 letters 2 or 3 syllables	*banlem* *terrow*	*esbort* *kisonic*	δδƱƱƘΦ δδƱƱƘΦ	δƧƔΞƔ ƔƧƔΞƔ
6	7 or 8 letters 2 or 3 syllables	7 or 8 letters 2 or 3 syllables	*opunate* *fundacy*	*displert* *pomable*	ШƔΞΞΦƽ ШƔΞΞΦƽ	ΨШΘЖƘΞ ΨƘΘŒƘΞ
7	8 or 9 letters	8 or 9 letters	*uncament*	*bepidity*	ЖϟƽΞƽƕ	ΔѲЖӘŒ

	3 or 4 syllables	3 or 4 syllables	*nanculous*	*delibance*	ЖӋƽΞƽƕ	ΔΘΔƏŒ

Twelve neurologically normal adults were asked to make syllable counting judgments on 350 randomly generated items (50 per level), emphasizing accuracy over speed. Accuracy was 94.4% ± 6.5% (range 78.6–98.6%). Most of the incorrect responses represented genuine errors. For just 9 out of 350 items (2.6%), less than two thirds of responses were correct. These items were rechecked and 9 pseudowords were deemed ambiguous or problematic. These words were included in the present imaging experiment, but will be excluded in future applications.

Adaptive semantic matching paradigm

The semantic paradigm has been described in detail previously (Wilson et al., 2018). In brief, two words were presented on the screen in the experimental condition. Participants pressed a button if the words were semantically related and did nothing otherwise. The perceptual control task was the same as that used in the other paradigms. Difficulty was manipulated by varying lexical frequency, concreteness, degree of relatedness, word length, age of acquisition, and presentation rate. Note that the data for the 16 participants on the semantic paradigm in the present study have been previously reported in Wilson et al. (2018), section 3.6, however the analyses were slightly different.

Implementation

The three paradigms were implemented in a MATLAB program called *AdaptiveLanguageMapping* (Wilson et al., 2018) using the Psychophysics Toolbox version 3 (Brainard, 1997; Pelli, 1997). *AdaptiveLanguageMapping* is freely available for download at *http://aphasialab.org/alm*. Deidentified imaging data will be made available on request.

Training

Each participant was trained in two phases. In the first phase, the researcher explained the tasks and manually presented examples of match and mismatch items in each condition. In the second phase, participants practiced each paradigm with the real experiment timing (except that the presentation rate was not yoked across conditions) to familiarize them with the pace of the experiment. Items presented in training were not repeated in the scanner. The difficulty levels achieved on each condition at the end of the practice session were used as the initial difficulty levels during the scanning session.

Patients with language impairments will benefit from further training in the scanner (Wilson et al., 2018), but this was not necessary for the neurologically normal participants in the present study.

Neuroimaging

Participants were scanned on a Philips Achieva 3T scanner with a 32-channel head coil at the Vanderbilt University Institute of Imaging Science. Visual stimuli were projected onto a screen at the end of the bore, which participants viewed through a mirror mounted to the head coil. Three functional runs of T2*-weighted BOLD echo planar images—one for each paradigm—were collected with the following parameters: 200 volumes + 4 initial volumes discarded; 35 axial slices in interleaved order; slice thickness = 3.0 mm with 0.5 mm gap; field of view = 220 × 220 mm; matrix = 96 × 96; repetition time (TR) = 2000 ms; echo time (TE) = 30 ms; flip angle = 75°; SENSE factor = 2; voxel size = 2.3 × 2.3 × 3.5 mm. The order of the three paradigms (rhyme, syllables, semantic) was counterbalanced across participants. Three-dimensional T1-weighted and coplanar T2-weighted structural images were acquired for intrasubject and intersubject registration. Participants responded using a button box held in their left hand.

Analysis of fMRI data

The functional data were first preprocessed with tools from AFNI (Cox, 1996). Head motion was corrected, with six translation and rotation parameters saved for use as covariates. Next, the data were detrended with a Legendre polynomial up to and including degree 2, and smoothed with a Gaussian kernel (FWHM = 6 mm). Then, independent component analysis (ICA) was performed using the fsl tool *melodic* (Beckmann & Smith, 2004). Noise components were manually identified with reference to the criteria of Kelly et al. (2010) and removed using *fsl_regfilt*.

The adaptive paradigms were modeled with boxcar functions encoding the block design; note that most blocks included correct and incorrect trials. These models were convolved with a hemodynamic response function (HRF) based on the difference of two gamma density functions (time to first peak = 5.4 s, FWHM = 5.2 s; time to second peak = 15 s; FWHM = 10 s; coefficient of second gamma density = 0.09), and fit to the data with the program *fmrilm* from the FMRISTAT package (Worsley et al., 2002). The six motion parameters were included as covariates, as were time-series from white matter and CSF regions (means of voxels segmented as white matter or CSF in the vicinity of the lateral ventricles) to account for nonspecific global fluctuations, and three cubic spline temporal trends.

The T1-weighted anatomical images were warped to MNI space using unified segmentation in SPM5 (Ashburner & Friston, 2005). Functional images were coregistered with structural images via coplanar T2-weighted structural images using SPM, and warped to MNI space.

Contrasts were created to compare each language condition to its perceptual control condition. Second level analyses were based on these contrasts, and also on contrasts between the paradigms to identify brain regions that were differentially recruited by different language tasks, e.g. (Rhyme – Perceptual_R) – (Semantic – Perceptual_S), where Perceptual_R refers to the perceptual task during the rhyme run, and Perceptual_S the perceptual task during the semantic run. All second level contrasts were thresholded at voxelwise $p < .001$, then corrected for multiple comparisons at $p < .05$ based on cluster extent using permutation testing implemented with the FSL tool *randomise* (Winkler et al., 2014). Specifically, null distributions were created by randomly inverting the signs of the contrast images (Nichols & Holmes, 2002).

Measures of validity

Sensitivity

One indicator of paradigm validity is the ability to activate, in individual participants, known phonological encoding regions in the left supramarginal gyrus and left ventral precentral gyrus. For completeness, we also quantified activation of the language regions in the left inferior frontal gyrus and left posterior temporal cortex.

Four ROIs were defined as follows. The parietal ROI was defined as the supramarginal gyrus (AAL region 63; Tzourio-Mazoyer et al., 2002) and inferior parietal lobule (AAL region 61). The ventral precentral ROI was defined as the precentral gyrus (AAL region 1), with a dorsal cutoff of $z < 40$. The frontal ROI was defined as the inferior frontal gyrus (AAL regions 11, 13 and 15), while the temporal ROI was defined as the middle temporal gyrus (AAL region 85), angular gyrus (AAL region 65), and the ventral part of the superior temporal gyrus (AAL region 81); specifically, voxels within 8 mm of the middle temporal gyrus. For all analyses, a gray matter mask was applied, obtained by smoothing the segmented gray matter proportion image with a 4 mm FWHM Gaussian kernel, then applying a cutoff of 0.25.

We quantified sensitivity in these ROIs in two ways: extent of activation and proportion of participants for whom activation exceeded 2,000 mm^3. Relative thresholds were used in each individual such that the top 5% of all gray matter voxels were considered active (Gross & Binder, 2014), and a cluster volume cutoff of 2,000 mm^3 was applied (Wilson et al., 2017, 2018).

Lateralization

Another indicator of validity is the extent to which activations in these regions are left-lateralized, given that language function is known to be left-lateralized in the majority of neurologically normal individuals. We therefore defined homotopic ROIs in the right hemisphere, and then calculated a laterality index (LI) for each paradigm and ROI using the standard formula $\text{LI} = (V_{\text{Left}} - V_{\text{Right}}) / (V_{\text{Left}} + V_{\text{Right}})$, where V_{Left} is the number of voxels activated in the left hemisphere, and V_{Right} is the number of voxels activated in the right hemisphere. LI ranges from –1 (all activation in the right hemisphere) to +1 (all activation in the left hemisphere).

To compare the degree of lateralization across paradigms, it was not possible to use parametric tests due to parietal and precentral regions not being activated in all participants, and also due to ceiling effects (LI = 1 reflecting complete lateralization). Therefore permutation testing was used. For each ROI, 10,000 permutations were carried out. The parietal and precentral ROIs were rarely activated by the semantic paradigm, so only the rhyme and syllable counting paradigms were compared. Each participant's two LIs (missing in several instances) were randomly assigned to the two conditions, and the test statistic was defined as the difference between the means of the non-missing values. The observed difference was compared to its null distribution. For the frontal and temporal ROIs, each participant's three LIs (missing in one instance) were randomly reassigned to the three conditions. The means of the non-missing values were then calculated for each condition. The test statistic was the largest difference between the three means.

Reliability

To estimate test-retest reproducibility, we split each run in half, refit the general linear model to each of the two halves independently, and then calculated the Dice coefficient of similarity (Rombouts et al., 1997) to assess the degree of activation overlap between the two resultant maps.

Analysis parameter sets

Analysis parameters such as voxelwise threshold and cluster volume cutoff have a strong impact on calculations of sensitivity, lateralization, and reliability (Wilke & Lidzba, 2007; Wilson et al., 2017, 2018). As stated above, we used an *a priori* set of parameters for the majority of our analyses, but we also explored the effect of changing these parameters. To ensure that our main findings were not dependent on threshold or cluster extent cutoff, we recalculated sensitivity and lateralization measures under seven different absolute thresholds ($p < .1$, $p < .05$, $p < .01$, $p < .005$, $p < .001$, $p < .0005$, and $p < .0001$), seven relative thresholds (top 10%, 7.5%, 5%, 4%, 3%, 2%, and 1% of most highly activated voxels), and four cluster volume cutoffs (none, 1,000 mm^3, 2,000 mm^3, 4,000 mm^3).

RESULTS

Behavioral results

The adaptive staircase procedure was intended to result in performance being similarly accurate in each condition. This was partially, but not entirely, successful, as evidenced by significant differences in accuracy by paradigm and condition (Fig. 2.1). In a repeated measures ANOVA, there was a significant interaction of paradigm by condition ($F(2, 30) = 19.18$, $p < .0001$). Follow-up tests paired *t*-tests showed that accuracy was lower on the rhyme paradigm than the semantic paradigm ($|t(15)| = 5.75$, $p < .0001$), and lower on the syllables paradigm than the semantic paradigm ($|t(15)| = 4.86$, $p = .0002$), while the two phonological paradigms did not differ from one another ($|t(15)| = 0.72$, $p = .48$). When each language paradigm was compared to its perceptual control condition, participants were less accurate on the phonological paradigms than on their control conditions (rhyme: $|t(15)| = 5.61$, $p < .0001$; syllables: $|t(15)| = 3.51$, $p = .0032$), but more accurate on the semantic paradigm than on its control condition ($|t(15)| = 2.19$, $p = .045$).

The mean difficulty level at which items were presented also showed a significant interaction of paradigm by condition ($F(2, 30) = 21.86$, $p < .0001$). Follow-up paired *t*-tests showed that difficulty levels were higher on the semantic paradigm than on the rhyme ($|t(15)| = 10.43$, $p < .0001$) or syllables paradigm ($|t(15)| = 4.66$, $p = .0003$). Difficulty levels were also higher on the syllables paradigm than on the rhyme paradigm ($|t(15)| = 3.59$, $p = .0027$). When each language paradigm was compared to its perceptual control condition, items were presented at lower difficulty levels on the phonological paradigms than on their control conditions (rhyme: $|t(15)| = 6.21$, $p < .0001$; syllables: $|t(15)| = 2.85$, $p = .0122$), but higher difficulty levels on the semantic condition compared to its control condition ($|t(15)| = 3.30$, $p = .0048$).

Finally, reaction times too showed a significant interaction of paradigm by condition ($F(2, 30) = 32.14$, $p < .0001$). Follow-up paired *t*-tests showed that reaction times were faster on the semantic paradigm than on the rhyme paradigm ($|t(15)| = 5.14$, $p < .0001$) or the syllables paradigm ($|t(15)| = 8.16$, $p < .0001$), and reaction times were faster on the rhyme paradigm than the syllables paradigm ($|t(15)| = 4.19$, $p = .0008$). Participants responded more quickly to the rhyme condition than to its control condition ($|t(15)| = 2.25$, $p = .040$), and more quickly to the semantic task than its control condition ($|t(15)| = 7.82$, $p < .0001$), while the syllables task did not differ from its control condition ($|t(15)| = 1.49$, $p = .16$).

Taken together, these findings indicate that the adaptive staircase procedure was only partially successful in matching performance across paradigms and conditions. Limitations following from this will be addressed in the discussion section.

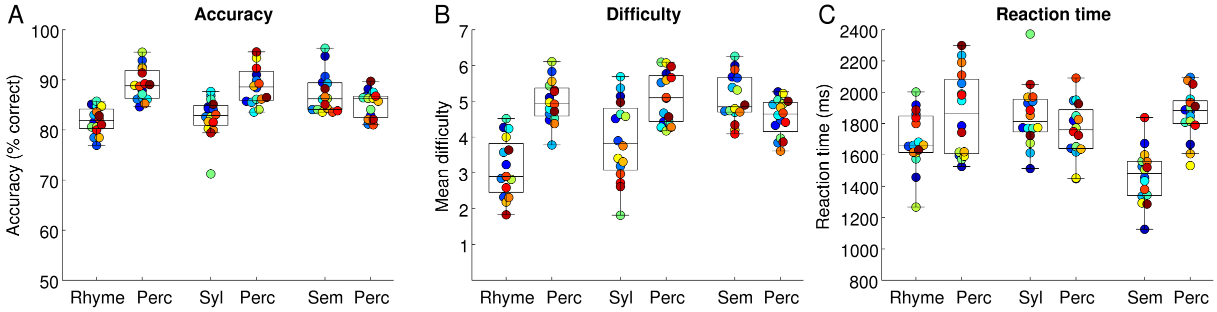


Figure 2.1 Behavioral results. Each participant is denoted with a unique color. (A) Accuracy. (B) Difficulty level of items presented. (C) Reaction time. Rhyme = Rhyme judgment; Syl = Syllable counting; Sem = Semantic matching; Perc = Perceptual control task.

Neuroimaging results

In group contrast maps (Fig 2.2A-C), all three paradigms activated left inferior frontal and left posterior temporal regions, but there were striking differences between paradigms in the inferior parietal and ventral precentral regions that have been associated with phonological encoding. These regions were activated by the rhyme judgment task (Fig 2.2A, green circles) and the syllable counting task (Fig 2.2B, blue circles), but not by the semantic task (Fig 2.2C). Direct contrasts between the three paradigms (Fig 2.2D-F) confirmed that these differences were statistically significant (Fig 2.2D, 2E).

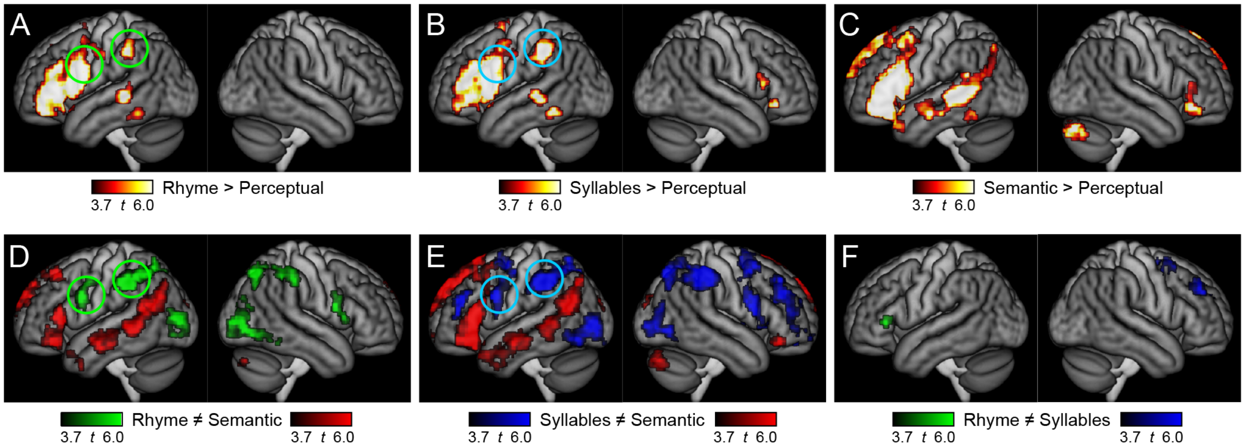


Figure 2.2 Activation maps for group contrasts. The top row shows each language task compared to its perceptual control task: (A) Rhyme > Perceptual; (B) Syllables > Perceptual; (C) Semantic > Perceptual. The bottom row shows pairwise contrasts between the language paradigms: (D) Rhyme ≠ Semantic; (E) Syllables ≠ Semantic; (F) Rhyme ≠ Syllables.

Coordinates and statistical details for all activated regions are provided in Table 3.

Table 2.3 Functional activations for group contrasts

Brain region	MNI coordinates			Extent (mm^3)	Max *t*	*p*
	x	y	z			
Rhyme						
Left IFG, pars opercularis, triangularis and orbitalis; anterior insula; ventral precentral gyrus	–46	20	14	40,304	12.16	0.0002
Left medial superior frontal gyrus	–6	24	47	3,192	6.18	0.0102
Left posterior STS	–61	–36	1	3,024	10.90	0.011
Left supramarginal gyrus	–48	–40	43	2,768	7.33	0.011
Left hippocampus	–17	–12	–15	1,424	6.09	0.013
Left supplementary motor area	–4	2	64	1,384	6.31	0.033
Left posterior inferior temporal gyrus	–50	–47	–17	1,240	5.40	0.038
Syllables						
Left IFG, pars opercularis, triangularis and orbitalis; anterior insula; ventral precentral gyrus	–46	19	16	40,032	12.85	0.0004
Left medial superior frontal gyrus	–3	12	54	9,256	10.47	0.0029
Left supramarginal gyrus	–50	–39	41	7,112	9.52	0.0035
Left posterior STS	–65	–34	–1	1,704	7.93	0.025
Right anterior insula	37	29	–4	1,296	7.20	0.038
Left inferior temporal gyrus	–50	–49	–15	1,136	6.49	0.046
Right IFG pars opercularis	52	17	15	1,128	5.58	0.046
Semantic						
Left IFG, pars opercularis, triangularis and orbitalis; anterior insula	–45	23	9	36,856	11.39	0.0010

Left STS, MTG and angular gyrus, posterior ITG	–56	–40	2	21,864	12.17	0.0015
Left medial superior frontal gyrus	–7	36	45	16,240	10.40	0.0018
Right cerebellum	23	–78	–32	9,328	12.44	0.0031
Left hippocampus	–20	–12	–14	4,024	8.61	0.0061
Right IFG, pars orbitalis	46	32	–7	3,464	7.37	0.0075
Right anterior calcarine sulcus	29	–48	13	1,808	5.85	0.017
Left anterior calcarine sulcus	–9	–52	9	1,184	6.00	0.036
Rhyme vs Semantic						
Right supramarginal gyrus, intraparietal sulcus, precuneus	33	–53	46	16,744	8.50	0.0005
Left supramarginal gyrus, intraparietal sulcus, precuneus	–36	–50	44	16,224	13.78	0.0005
Right inferior temporal gyrus, middle occipital gyrus	43	–77	1	8,264	8.50	0.0018
Left middle occipital gyrus	–38	–86	4	5,240	7.55	0.0033
Right pars opercularis, ventral precentral gyrus	48	11	15	3,384	6.72	0.0055
Left ventral precentral gyrus	–50	3	30	2,888	7.76	0.0068
Syllables vs Semantic						
Right supramarginal gyrus, intraparietal sulcus, precuneus	36	–57	38	37,680	14.65	0.0001
Left supramarginal gyrus, intraparietal sulcus, precuneus; middle occipital gyrus	–38	–57	32	30,736	14.90	0.0001
Left supplementary motor area, medial superior frontal gyrus, precentral sulcus	–14	1	55	9,608	7.66	0.0016
Right middle frontal gyrus	38	40	26	8,640	6.77	0.0021
Right precentral sulcus	31	3	57	8,128	6.38	0.0022
Right ventral precentral gyrus; anterior insula	45	11	16	6,640	7.08	0.0029
Left precentral gyrus	–52	4	30	3,536	9.60	0.0068
Left middle frontal gyrus	–39	40	23	2,600	8.74	0.011
Right orbital frontal cortex	24	53	–10	1,416	5.73	0.027
Rhyme vs Syllables						
Left pars triangularis	–48	29	5	1,144	6.07	0.036
Semantic vs Rhyme						
Left STS, MTG and angular gyrus	–54	–45	7	20,272	7.96	0.0003
Left precuneus, posterior cingulate gyrus	–9	–51	19	8,896	7.36	0.0010
Left medial superior frontal gyrus	–12	48	36	7,920	7.12	0.0014
Left IFG, pars triangularis and orbitalis	–49	29	–4	5,760	7.98	0.0023
Bilateral cuneus	0	–86	22	2,720	5.98	0.0079
Right cerebellum	25	–81	–33	1,152	5.50	0.038
Left temporal pole	–44	11	–31	1,032	5.29	0.046
Semantic vs Syllables						
Left medial superior frontal gyrus, superior frontal sulcus	–10	43	36	22,832	9.08	0.0003
Left anterior STS, MTG; IFG pars triangularis and orbitalis	–50	15	–8	21,992	11.23	0.0003
Left posterior STS, MTG, angular gyrus	–52	–57	16	16,152	8.95	0.046
Left precuneus, bilateral anterior calcarine sulcus	–4	–54	16	10,552	10.59	0.0011
Bilateral cuneus	0	–86	20	7,584	8.48	0.0024
Right cerebellum	25	–80	–32	5,776	7.05	0.0034
Left parahippocampal gyrus	–27	–36	–16	1,320	5.87	0.031
Right IFG pars orbitalis	41	33	–14	1,160	7.64	0.038
Syllables vs Rhyme						
Right middle frontal gyrus	33	37	34	3,160	6.82	0.0028
Right superior frontal gyrus	26	6	57	2,576	5.99	0.0056

MNI coordinates indicate centers of mass. IFG = inferior frontal gyrus; STS = superior temporal sulcus; MTG = middle temporal gyrus; ITG = inferior temporal gyrus

Sensitivity

Sensitivity was compared across paradigms and ROIs (Fig. 2.3). In the left supramarginal ROI, the extent of activation differed across the three paradigms ($F(2, 30) = 20.86$, $p < .0001$), with more extensive activation for the rhyme paradigm (4.92 ± 3.67 cm^3) than the semantic paradigm (1.02 ± 1.16 cm^3, $|t(15)| = 5.05$, $p < .0001$) and more extensive activation for the syllable counting paradigm (5.09 ± 3.31 cm^3) than the semantic paradigm ($|t(15)| = 6.50$, $p < .0001$), while the phonological paradigms did not differ from one another ($|t(15)| = 0.23$, $p = .82$). On both phonological paradigms, most participants showed supramarginal activations of at least 2000 mm^3 extent: 13 out of 16 on the rhyme paradigm and 13 out of 16 on the syllable counting. In contrast, on the semantic paradigm only 2 out of 16 participants showed such activation. This difference was statistically significant ($\chi^2(2) = 20.74$, $p < .0001$).

In the left ventral precentral ROI, the extent of activation differed across the three paradigms ($F(2, 30) = 24.81$, $p < .0001$), with more extensive activation for the rhyme paradigm (3.56 ± 1.46 cm^3) than the semantic paradigm (1.18 ± 0.94 cm^3, $|t(15)| = 7.79$, $p < .0001$) and more extensive activation for the syllable counting paradigm (3.52 ± 1.22 cm^3) than the semantic paradigm ($|t(15)| = 8.26$, $p < .0001$), while the phonological paradigms again did not differ ($|t(15)| = 0.20$, $p = .85$). On both phonological paradigms, most participants showed inferior parietal activations of at least 2000 mm^3 extent: 14 out of 16 on the rhyme paradigm and 15 out of 16 on the syllable counting. In contrast, on the semantic paradigm only 3 out of 16 participants showed such activation. This difference was statistically significant ($\chi^2(2) = 24.94$, $p < .0001$).

All three paradigms activated the inferior frontal gyrus ROI. There were modest but significant differences in extent (*F*(2, 30) = 5.57, *p* = .015), with more activation for the semantic paradigm (17.85 ± 3.47 cm^3) than the syllables paradigm (14.62 ± 4.26 cm^3, |*t*(15)| = 2.69, *p* = .017) and more activation for the rhyme paradigm (17.22 ± 3.11 cm^3) than the syllables paradigm (|*t*(15)| = 3.57, *p* = .0028); the semantic and rhyme paradigms did not differ from one another (|*t*(15)| = 0.58, *p* = .57).

All three paradigms also activated the posterior temporal ROI, again with significant differences in extent (*F*(2, 30) = 50.78, *p* < .0001). In this case, the semantic paradigm produced much more extensive activation (15.76 ± 4.75 mm^3) than the rhyme paradigm (7.45 ± 3.55 cm^3, |*t*(15)| = 7.56, *p* < .0001) or the syllables paradigm (6.36 ±2.80 cm^3, |*t*(15)| = 7.32, *p* < .0001), which did not significantly differ from one another (|*t*(15)| = 2.13, *p* = .051).

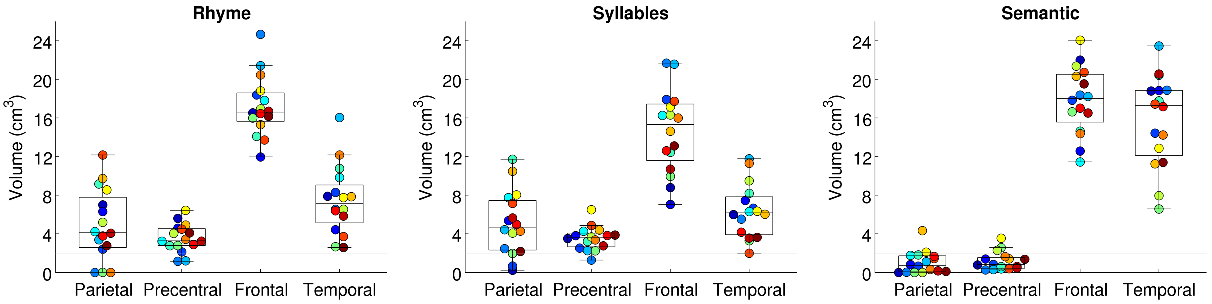


Figure 2.3 Extent of activation for each paradigm in each region of interest. The gray horizontal line indicates the 2,000 mm^3 extent of activation which was our threshold for considering a region activated. Each participant is denoted with a unique color, matching Fig. 2.1.

Lateralization

Lateralization was compared across paradigms and ROIs (Fig. 2.4). In the supramarginal gyrus, the rhyme paradigm yielded more left-lateralized activations (LI = 0.99 ± 0.03) than the syllables paradigm (LI = 0.71 ± 0.38, permutation test, *p* = .0179). In the ventral precentral gyrus, the difference in lateralization between these paradigms was not significant (rhyme: LI = 0.84 ± 0.24; syllables LI = 0.74 ± 0.32; *p* = .1803). Note that lateralization was not assessed for the semantic paradigm in these regions because they were activated in so few participants. The frontal and temporal ROIs did not show significant lateralization differences between the three paradigms (frontal: omnibus *p* = .1213; temporal: omnibus *p* = .0646).

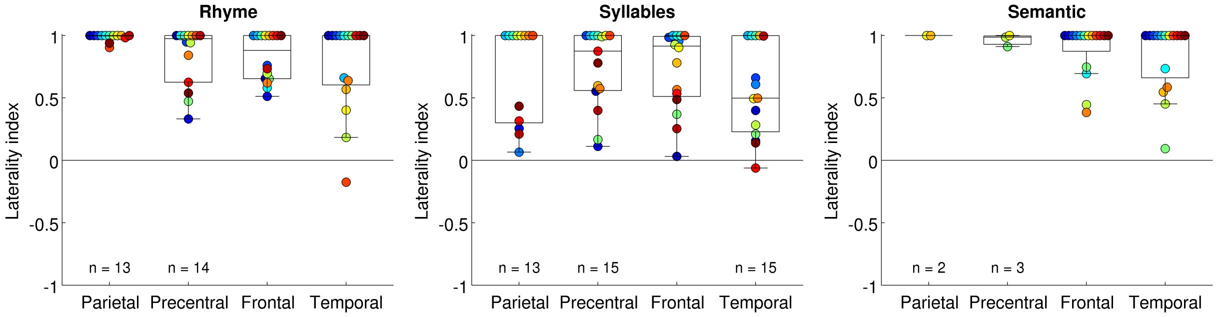


Figure 2.4 Lateralization of activation for each paradigm in each region of interest. Each participant is denoted with a unique color, matching previous figures.

Reliability

In the split-half analyses of test-retest reproducibility (Fig 2.5), the Dice coefficient of similarity differed across the three paradigms ($F(1, 30) = 6.57$, $p = .0057$). The rhyme paradigm (Dice = 0.61 ± 0.02) and semantic paradigm (Dice = 0.66 ± 0.03) both showed better split-half reproducibility than the syllables paradigm (Dice = 0.52 ± 0.03; rhyme: $|t(15)| = 2.46$, $p = .026$; semantic: $|t(15)| = 3.13$, $p = .0068$), but the rhyme and semantic paradigms did not differ from one another ($|t(15)| = 1.29$, $p = .22$).

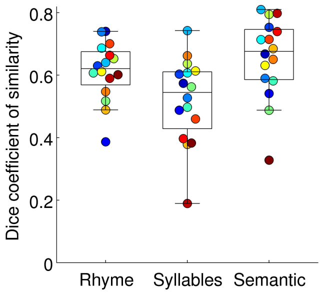


Figure 2.5 Dice coefficients of similarity, based on split-half analyses, indicating test-retest reproducibility of each paradigm. Each participant is denoted with a unique color, matching previous figures.

Effect of analysis parameters

The impact of varying voxelwise thresholds and cluster size cutoffs on sensitivity, lateralization, and reliability of activations is shown in Fig. 2.7. These analyses showed that the main patterns described above held regardless of the voxelwise threshold or the cluster size cutoff. In particular, the supramarginal gyrus and ventral precentral gyrus were activated by both phonological paradigms under a wide range of analysis parameters, whereas these regions were activated by the semantic paradigm only under very liberal thresholds, in which case activations appeared to be mostly extensions of adjacent semantic regions (angular gyrus, inferior frontal gyrus). Phonological activations in these regions were also left-lateralized under a wide range of analysis parameters, generally more so for the rhyme paradigm. The rhyme and semantic paradigms were generally comparable in reliability, while the syllables paradigm was somewhat less reliable.

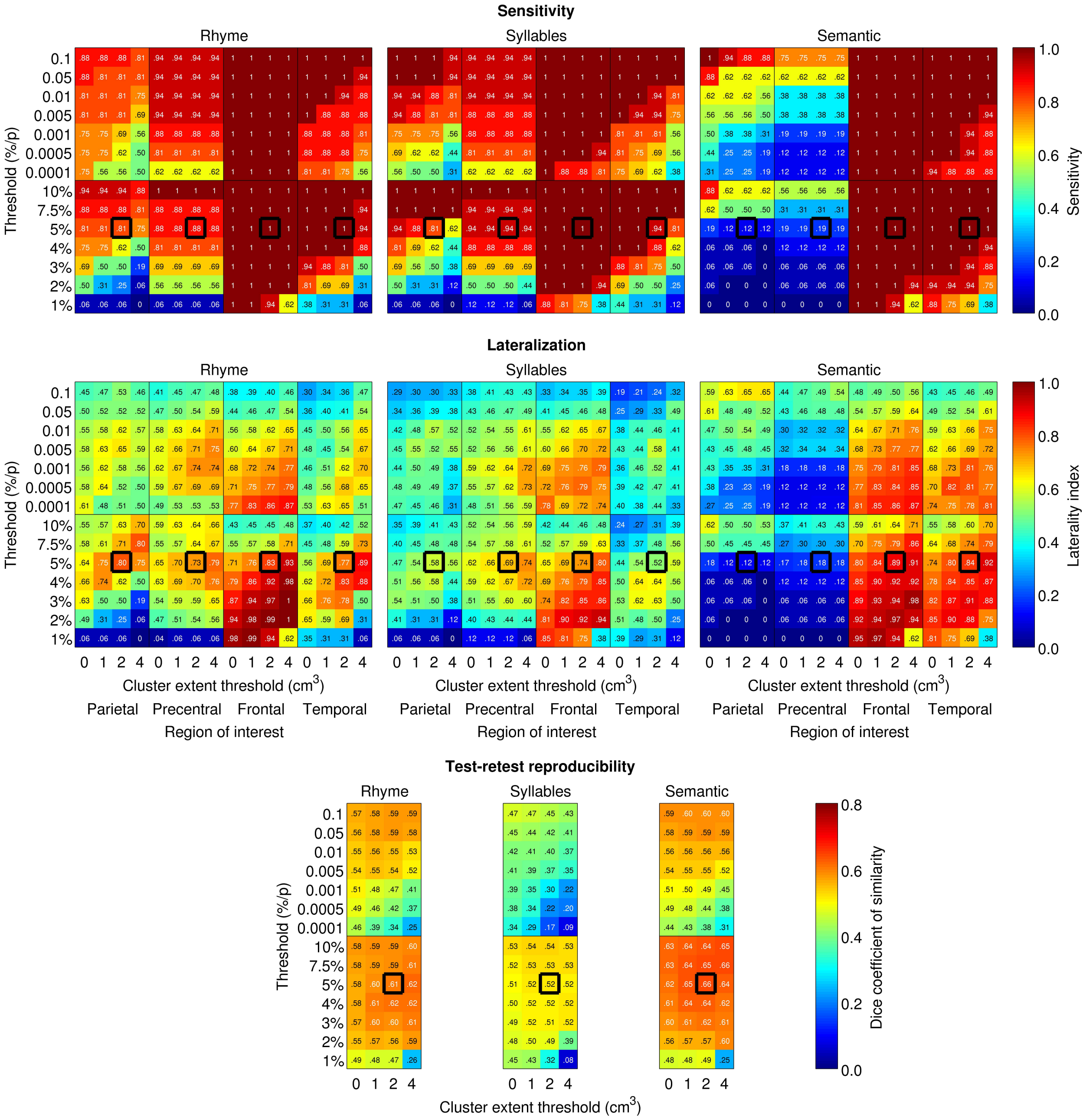


Figure 2.6 Impact of analysis parameters on sensitivity, laterality, and reliability. Thick black outlines denote the *a priori* analysis parameters. Sensitivity for detection, lateralization indices, and Dice coefficients are plotted as a function of absolute and relative voxelwise thresholds (*y* axes), region of interest (*x* axes) and minimum cluster volume (*x* axes). Detection was defined as activation of at least 2,000 mm^3 in the relevant ROI.

DISCUSSION

We found that the left supramarginal and left ventral precentral regions previously identified as the neural substrates of phonological encoding were robustly activated in most individual participants by the adaptive rhyme judgment and syllable counting paradigms, supporting the validity of these paradigms for mapping phonological regions. These activations were generally left-lateralized, especially for the rhyme paradigm. [The rhyme paradigm demonstrated good reliability (Dice coefficient = 0.61), and was significantly more reliable than the syllables paradigm. –take this part out if we're not talking about the split half stuff] For these reasons, we advocate the rhyme paradigm for future applications of this work. In contrast, our previously described semantic matching paradigm, which reliably activates inferior frontal and posterior temporal language regions, rarely yielded activations in the supramarginal gyrus or ventral precentral gyrus.

Left-lateralized inferior frontal and posterior temporal language regions were robustly activated by all three paradigms. The left inferior frontal activation was roughly similar across paradigms, while the temporal activation was considerably more extensive for the semantic paradigm, which is not surprising given the localization of the semantic network (Binder et al., 2009). Taken together, the findings from the three paradigms suggest that the left inferior frontal region, and a relatively circumscribed region centered on the left superior temporal sulcus, are involved in language processing in general (Knecht et al., 2003; Seghier et al., 2011; Springer et al., 1999; Tzourio-Mazoyer et al., 2010; Bradshaw et al., 2017; Wilson et al., 2018), while the supramarginal gyrus and ventral precentral gyrus are specifically driven by phonological encoding (Gitelman et al., 2005; McDermott et al., 2003; Price et al., 1997), and more extensive temporal lobe regions are involved in semantic processing (Binder et al., 2009). This circuitry is well established at the population level; what the present study adds is the potential to map these regions with distinct functions in individual participants.

Rhyme judgment paradigms have been implemented in many previous studies (e.g. Baciu et al., 2001, 2005; Billingsley et al., 2001; Booth et al., 2002, 2006; Burton et al., 2005; Clements et al., 2006; Cousin et al., 2007; Gitelman et al., 2005; Lurito et al., 2000; McDermott et al., 2003; Morrison et al., 2016; Pillai & Zacà, 2011; Poldrack et al., 2001; Pugh et al., 1996; Seghier et al., 2004; Shaywitz et al., 1995; Zacà et al., 2012), and syllable counting paradigms have been used in a few studies (DeMarco et al., 2016; Poldrack et al., 1999; Price et al., 1997; Trojano et al., 2000). Rhyme judgment paradigms have been shown to be highly lateralizing, making them an excellent choice for presurgical language mapping (Baciu et al., 2001, 2005; Clements et al., 2006; Cousin et al., 2007; Lurito et al., 2000; Pillai & Zacà, 2011; Zacà et al., 2012; Morrison et al., 2016; see Bradshaw et al., 2017 for review), while syllable counting paradigms appear to yield somewhat more bilateral activation patterns (e.g. Price et al., 1997; Trojano et al., 2000), like we found in the present study. Most previous group studies using either of these tasks have not reported activation of all four regions that we studied: the left supramarginal gyrus, left ventral precentral gyrus, left inferior frontal gyrus, and left posterior temporal cortex. To our knowledge, only one previous study has explicitly reported sensitivity to detect activation in putative phonological regions: Seghier et al. (2004) found that for a rhyme judgment task, the inferior parietal lobule was activated in 13/26 participants and the precentral gyrus in 16/26 participants. The core language areas were more consistently activated: the inferior frontal gyrus in 23/26, and the superior or middle temporal gyrus in 19/26. There are many factors that may account for the apparent greater sensitivity of our paradigms, including number of blocks (we used 10 blocks per condition while Seghier and colleagues used 5 blocks per condition), our use of pseudowords rather than real words (Poldrack et al.,1999), and the adaptive staircase procedure, which ensured that language tasks as well as the control tasks were challenging at all times, requiring each participant to engage in focused and highly constrained linguistic and cognitive processing.

Besides potentially increased sensitivity, another advantage of using adaptive designs is that the paradigms are likely to be more feasible for individuals with aphasia. This may improve the accuracy of language mapping in presurgical patients with language deficits due to tumors or epileptogenic foci in or around language regions. It also makes it feasible to study the functional reorganization of phonological encoding regions in recovery from aphasia. A major challenge in studies of neuroplasticity has been designing language tasks that individuals with aphasia are able to perform (Price et al., 2006; Geranmayeh et al., 2014). Our adaptive paradigms have the potential to equate performance between patients and controls, as well as across the same individual at multiple time points over the course of recovery.

Limitations

Our study had several noteworthy limitations. First, phonological encoding regions were not identified in every participant. The rhyme paradigm, which we advocate for future use, activated the left supramarginal gyrus in 13 out of 16 participants, and the ventral precentral gyrus in 14 out of 16. In contrast, the core inferior frontal and posterior temporal language regions were activated with 100% sensitivity by the rhyme and semantic paradigms (see also Wilson et al., 2018), and in all but one participant by the syllable counting paradigm. Seghier et al. (2004) reported the same of pattern of results whereby phonological regions were less robustly activated than core regions.

Second, we were interested in identifying the left-lateralized fronto-parietal regions that are thought to be involved in the phonological encoding stage of speech production, since these regions are critical for speech production yet are not well localized by other language mapping paradigms. However, while our phonological contrasts certainly implicated phonological encoding, they also implicated orthographic decoding, as well as metalinguistic operations on the outputs of phonological encoding (rhyme judgment, syllable counting). It is likely that these other decoding and metalinguistic processes involved in our tasks, which are generally phonological in nature, have rather similar neural substrates to phonological encoding specifically (Booth et al., 2002; Jobard et al., 2003; Mechelli et al., 2003; Taylor et al., 2013; Wilson et al., 2009). However our study did not allow us to unequivocally dissociate phonological encoding from other types of phonological processing.

Third, our attempt to match accuracy and reaction time across the language and perceptual tasks, and across the three different language paradigms, was only partially successful. At least four possible reasons can be identified for the failure to obtain perfectly matched behavioral results. First, there were some ceiling effects, that is, some participants made consecutive correct responses at the highest level, and thus would have been forced to perform at even higher levels if there had been any higher levels. Second, some items were ambiguous, meaning that difficulty may sometimes have been adjusted up or down when the opposite should have occurred. These items have been excluded from the paradigms that we have made available for future applications. Third, presentation rate had to be yoked across the language and perceptual conditions, in order to avoid even more serious confounds of mismatched sensorimotor demands. Fourth, string length was matched across the language and perceptual conditions to avoid visual confounds. The matching of presentation rate and string length created contingencies between the conditions that prevented them from adapting completely independently.

Fortunately, the impact of the imperfect matching was considerably mitigated by the fact that the language conditions had shorter reaction times than the perceptual condition for the rhyme and semantic paradigms, and equivalent reaction times for the syllable counting paradigm. This is important because it means that domain-general regions (Fedorenko et al., 2013), which show increased signal with increased time on task (Binder et al., 2005), cannot be misidentified as language regions. However, despite reaction times being shorter, participants were less accurate on the rhyme and syllable counting conditions than on their perceptual control conditions. This means that activations for these contrasts may include components related to the commission or monitoring of errors. However, error monitoring has been associated with medial frontal regions (Ito et al., 2003), whereas the phonological regions have not generally been linked to error monitoring. Moreover, these regions did not show positive correlations between error rate differences and activation extent. These considerations suggest that the supramarginal and ventral precentral regions we identified are unlikely to represent confounds of accuracy.

Finally, while this study has provided considerable evidence to establish the validity and reliability of our phonological paradigms, we have not yet investigated their feasibility in individuals with aphasia (Wilson et al., 2018). In a lesion-symptom mapping study, Pillay et al. (2014) showed that many individuals with aphasia were able to perform above chance on a rhyme matching task using real words. This suggests that many patients should be able to perform our rhyme judgment paradigm, although our use of pseudowords, as well as the timing demands of functional imaging, are likely to make our task more challenging. DeMarco (2016) scanned eight individuals with aphasia on a syllable counting paradigm that was a precursor to the paradigm described here. All patients were able to attempt the task, however only four of the eight performed above chance. In our preliminary experiences with individuals with aphasia, it appears that patients are more likely to understand and perform above chance on rhyme judgment than syllable counting. This provides a third reason for preferring the rhyme judgment paradigm in future applications.

CHAPTER III

Experiment II: Application to aphasia

INTRODUCTION

The paradigms we developed are capable of activating the regions involved in language processing with great sensitivity even at the individual level: both the semantic match and rhyme tasks activate the core frontal and temporal areas, while the semantic paradigm additionally captures temporal semantic regions, and the rhyme paradigm captures precentral and supramarginal phonological regions.

With their effectiveness illustrated in neurotypical controls, the paradigms can then be applied to populations with brain damage. We can investigate the neural mechanisms underlying functional recovery by having patients with aphasia perform the tasks and analyzing the brain regions involved, across patients and over time. By examining the activation maps produced by the patients with aphasia and by the neurologically normal controls, we can identify any systematic functional changes that occur as a result of brain damage.

The work described in Chapter Two may subserve many lines of future research. The semantic and rhyme paradigms can be administered in isolation or in conjunction, in individuals with or without neurological conditions, and in patients in various stages of recovery. The present work will be centered on one out of a great breadth of potential applications: activation associated with phonological processing in a heterogeneous group of patients with aphasia. To gain a broad understanding of what these patterns look like, we include patients with various aphasia types at various post-stroke time points.

We will investigate the ability of our offline language assessments and in-scanner performance measures to reflect a patient's deficits by evaluating them against a subjective characterization of the patient's presentation. With respect to the neuroimaging data, we will first examine generalized differences between how damaged and undamaged language networks process linguistic stimuli by comparing the activation maps from the neurotypical and aphasic participants as groups. We will then explore the relationship between activation and better performance on behavioral measures to determine which areas are associated with superior language processing when the typical network is disturbed.

METHODS

Participants

Neurotypical controls

An additional sixteen neurologically normal controls were recruited from the Nashville area for a total of 32 controls when combined with the sixteen participants included so far. Recruitment of this group matched the patient group as much as possible in age, sex, handedness, and education (age 55.1 ± 16.6 years (range 23–83 years); 11 male, 21 female; 25 right-handed, 5 left-handed, 2 ambidextrous; education 16.7 ± 1.9). Inclusion criteria were: (1) fluent English speaker; (2) no history of neurological disease or speech/language or learning disabilities; (3) no contraindications for MRI. These healthy controls were scanned with a single fMRI session while performing the semantic and rhyme paradigms, in order to assess the paradigms' effectiveness in a greater sample size. Neuroimaging and preprocessing protocols were similar to those described in Chapter Two, except that all participants performed the tasks in the same order (semantic, then rhyme) and did not perform the syllable counting task. Language and cognitive function were evaluated with a single form of the QAB and the Mini-Mental State Examination.

Individuals with aphasia

36 patients with aphasia were recruited from an acute stroke database and chronic aphasia group at Vanderbilt Medical Center (age 57.3 ± 13.8 years (range 26–82 years); 18 male, 18 female; 34 right-handed, 2 left-handed; education 13.5 ± 2.5 years). Of these 36 patients, five were unable to perform the in-scanner rhyme task because of degree of aphasia severity or suspected low reading ability at baseline. As such, the remaining 31 patients (age 57.0 ± 13.1 years (range 34–82 years); 16 male, 15 female; 29 right-handed, 2 left-handed; education 13.6 ± 2.5 years) were included in the analyses involving activation maps and behavioral measures related to the rhyme task. Inclusion criteria for the acute patients were: (1) 18 - 90 years of age; (2) recruitment within 5 days of stroke; (3) lesion size of at least 0.5 cm^3 resulting in aphasia; (4) fluent English speaker; (5) no medical or neuropsychological conditions severe enough to affect study participation (e.g. transfer to hospice, impaired baseline from previous stroke).

Acute patients were scanned while performing the paradigms at multiple time points with the goal of collecting data at 1 month, 3 months, and 1 year post-stroke. Neuroimaging analysis was conducted on each patient's most recent scan, achieving a cross-sectional examination of the data (1 month: n = 5; 3 months: n = 16; 1 year: n = 11). Patients were also administered the extended Quick Aphasia Battery (Wilson et al., 2018), Pyramids and Palm Trees Test (Howard and Patterson, 1992), and a questionnaire about living with aphasia. Chronic patients (n = 4) completed the assessments and a single scanning session. Inclusion criteria for chronic patients consisted of: (1) 18 – 90 years of age; (2) documentation of a presence or history of aphasia of any non-progressive etiology; (3) at least 6 months post onset; (4) fluent and literate in English premorbidly; (5) no dementia, impaired cognitive or language function at baseline for any other reason, major psychiatric disorders, serious substance abuse or withdrawal, or any contraindication to MRI. Note that individuals with a history of multiple strokes were included, as long as their most recent stroke was at least 6 months prior to participation.

Within the 31 patients who successfully completed the rhyme task, the distribution of time points analyzed was similar to that of the full 36 patients (1 month: *n* = 4; 3 months: *n* = 13; 1 year: *n* = 11; chronic: *n* = 3).

Quantification of phonological processing

For an additional measure indexing phonological encoding, patients' connected speech samples were rated to reflect subjective appraisals of their phonological ability. This measure is meant to complement the objective language assessment measures from the QAB and PPT. An investigator who was not involved in the administration of the patients' speech evaluations, and was not cognizant of the lesion location of each patient while rating, made these judgments based on the 0-4 (least to most severe) perceptual rating scale described in Casilio et al. (2019).

Training

Patients were trained on the tasks before performing them in the scanner. Based on data collected in Wilson et al., (2018), we estimated that training would take approximately 5 to 15 minutes for each patient depending on severity of deficits. In practice, training took 3-10 minutes. This illustrates that most patients understand the concept of the task quickly even if their language deficits later prevent them from reaching high difficulty levels during the performance of the task. Audiovisual recordings were taken during the administration of this training and the behavioral assessments for the purpose of scoring and improvement of training protocols.

Analysis of fMRI data

Lesion masks for the patient images were created through manual demarcation based on T1-weighted and FLAIR images. First level models were fit independently for each of the functional runs (one per paradigm). The adaptive rhyme paradigm was modelled with a simple boxcar function, reflecting the block design.

Comparing the activation maps for the paradigm between the patient and control groups allows us to investigate any consistent patterns in functional localization following brain damage. We conducted a two-group between groups *t*-test in SPM5 with a voxelwise threshold of $p < 0.005$ and corrected for multiple comparisons at $p < 0.01$ based on cluster extent. We predicted less activation in the commonly lesioned language-related regions for the aphasia patients relative to the healthy control group. As a group, the individuals with aphasia were expected to show decreased activation in the frontal, temporal, and parietal areas established as having a role in language processing, due to the damage sustained that resulted in aphasia. Furthermore, in line with functional reorganization literature to date, we hypothesized that language function will primarily depend on spared or perilesional tissue, capitalizing on residual components of the original language network with little to no recruitment of the opposite hemisphere.

The activation maps produced by the rhyme paradigm should depict the regions involved in phonological processing, thereby illustrating whether and how these regions have changed as a result of damage. We correlated activation with the patients' behavioral measures by including each one as a covariate in a separate model. These measures consisted of performance on the in-scanner task (accuracy on the rhyme paradigm) and on the language assessment (overall QAB score, repetition subtest score, and phonological paraphasia score). We predicted that better performance on these measures would be associated with greater activation in the supramarginal and precentral gyri for patients with these regions intact. Moreover, for patients who have damage in these regions, better phonological performance may be correlated with activity in compensatory regions. In other words, if phonological processing has indeed functionally reorganized to other areas, activation in those areas should be associated with better phonemic encoding given that the original areas are compromised. We also correlated activation with the subjective perceptual ratings of each patient's phonological encoding ability.

Finally, we correlated activation with damage to the phonological regions. A parietally-damaged patient was defined as having at least 20% damage to the parietal ROI consisting of the supramarginal gyrus (AAL 63) and inferior parietal lobule (AAL region 61). We repeated the process with patients who did and did not have ventral precentral gyrus damage, with a ROI defined as the precentral gyrus (AAL 1) with a dorsal cutoff of $z < 40$.

RESULTS

Behavioral results

Repetition scores spanned a wide range in our sample (7.88 ± 2.05, range 2.5-10). Accuracy on the in-scanner rhyme task showed that, despite their deficits, patients tended to maintain the 80% accuracy that the paradigm was designed to adapt for (0.78 ± 0.054, range 0.67-0.88).

Correlation analyses were carried out between the perceptual ratings and behavioral measures in order to discern which one was most true to subjective assessments of the patients' abilities. One patient was missing accuracy data because of button box issues and another was missing a phonological paraphasia score because he produced too few words to viably gauge his speech sound processing. Corresponding scores for these two patients were excluded from analyses involving accuracy or phonological paraphasia score. Subjective ratings were extremely highly correlated with repetition score (r = -0.8848) and overall QAB score (r = -0.8917). Repetition scores and overall QAB scores were highly correlated with each other (r = 0.9494). Subjective ratings were moderately correlated with phonological paraphasia score (r = -0.4087) and showed little relationship with accuracy (r = -0.2174). Accuracy and repetition scores themselves were not associated with each other (r = 0.1020).

Neuroimaging results

Coordinates and statistical details for all significant clusters are given in Table 4.

Table 3.1 Activation clusters for functional contrasts

Brain region	MNI coordinates			Extent (mm^3)	Max t	p
	x	y	z			
Neurologically normal controls						
left IFG pars opercularis, pars triangularis, pars orbitalis, precentral gyrus, inferior parietal lobule, supramarginal gyrus, angular gyrus, middle temporal gyrus, inferior temporal gyrus,fusiform gyrus, anterior insula, basal ganglia, thalamus	-45	0	13	134536	14.82	0.000
left medial superior frontal gyrus, supplementary motor area	-3	19	52	16680	10.61	0.000
right IFG pars triangularis, pars orbitalis, pars opercularis	48	30	0	15824	7.01	0.000
right cerebellum	25	-70	-28	9032	8.05	0.000
right caudate	13	12	8	2416	4.45	0.034
Patients with aphasia						
left pars opercularis, pars triangularis, pars orbitalis, precentral gyrus, anterior insula	-46	20	19	37552	7.06	0.000
left medial superior frontal gyrus, supplementary motor area	-4	18	53	9920	8.32	0.000
left inferior temporal gyrus, fusiform gyrus	-54	-44	-11	9456	6.49	0.000
left supramarginal gyrus, angular gyrus	-41	-55	44	6640	4.36	0.003
right IFG pars orbitalis	43	42	-9	4304	4.82	0.001
right cerebellum	28	-67	-28	3576	5.55	0.000
Patients vs NNC						
bilateral precuneus	-1	-59	46	26360	4.86	0.000
left lingual gyrus	-15	-70	-10	8760	4.73	0.000
right lingual gyrus, fusiform gyrus	22	-60	0	7856	4.72	0.000
aphasia covariate with accuracy						
right IFG pars triangularis	44	32	7	2800	5.09	0.013
aphasia covariate with level						
left medial superior frontal gyrus, superior frontal gyrus	-9	35	41	3832	4.8	0.002
aphasia covariate with QAB						
right inferior temporal gyrus, fusiform gyrus, intraparietal sulcus, occipital lobe, cerebellum	32	-72	-1	40912	6.43	0.000
left inferior temporal gyrus, fusiform gyrus, occipital lobe, cerebellum	-40	-68	-15	17256	5.48	0.049
left superior parietal lobule, intraparietal sulcus	-23	-64	49	4064	4.67	0.000
right posterior superior frontal sulcus	34	-2	57	2976	5.13	0.009
right IFG pars triangularis	50	26	21	2560	4.85	0.021
right inferior frontal junction	47	12	32	2520	4.14	0.001
left posterior superior frontal sulcus	-23	2	49	2160	5.64	0.023
aphasia covariate with repetition						
right inferior temporal gyrus, fusiform gyrus, intraparietal sulcus, occipital lobe, cerebellum	33	-72	-3	38824	5.78	0.000
left inferior temporal gyrus, fusiform gyrus, occipital lobe	-40	-69	-15	18848	5.76	0.000
left superior parietal lobule, intraparietal sulcus	-23	-63	50	4032	4.76	0.001
right posterior superior frontal sulcus	34	-2	59	2488	4.73	0.025
aphasia covariate with phonemic paraphasias						
left inferior frontal sulcus	-29	40	15	3392	5.68	0.004
left IFG pars orbitalis	-30	45	-12	2744	5.2	0.015
aphasia covariate with repetition given lesion load						
left occipital lobe	-29	-77	33	4696	5.25	0.002
right angular gyrus, intraparietal sulcus	31	-61	49	4360	4.12	0.000
right supramarginal gyrus	48	-33	51	3680	5.36	0.004
right occpital lobe	38	-69	24	3448	4.33	0.001
right inferior temporal gyrus	51	-60	-20	2376	3.96	0.030

Group analyses

The group analysis for neurologically normal controls showed the rhyme task activated the frontal, midline, supramarginal, posterior and inferior temporal regions. Areas of greater activation for the perceptual control condition largely consisted of bilateral occipital cortex. These patterns were highly consistent with the typical language and visual processing networks respectively.

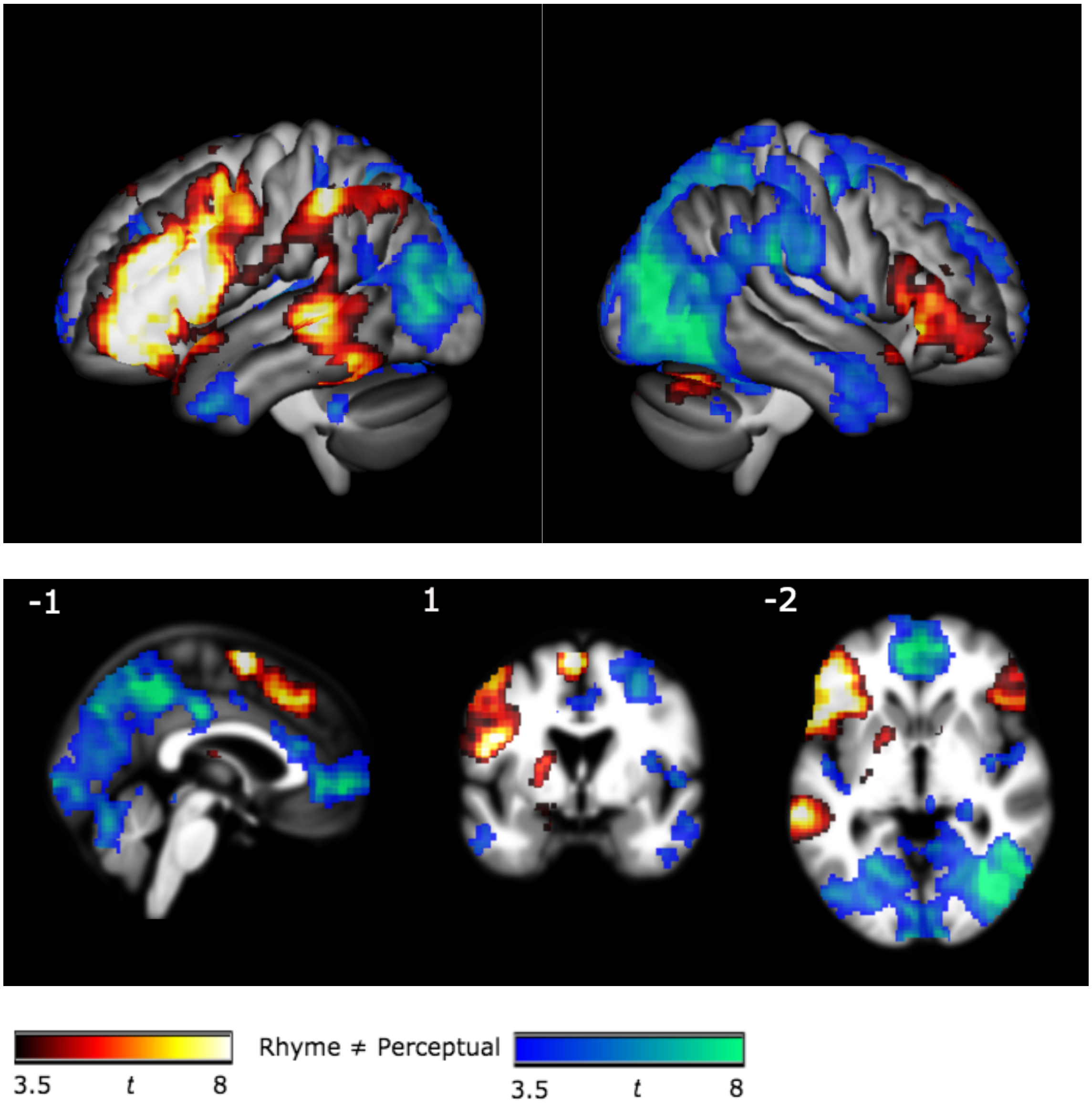


Figure 3.1 Group analysis for the rhyme vs. perceptual controls contrast in 32 neurologically normal controls. Both tails are shown: hot colors indicate areas of greater activation for the rhyme condition, cool colors reflect greater activation for the control condition.

The group analysis for patients demonstrated a pattern of activation similar to that of the neurotypical control group—clusters of activation in frontal, midline, supramarginal, and temporal regions. The patients' activated network generally appeared to be a less robust version of the controls' network, involving the same constellation of frontal, parietal, and temporal areas but with clusters that are lower in magnitude and more restricted in size.

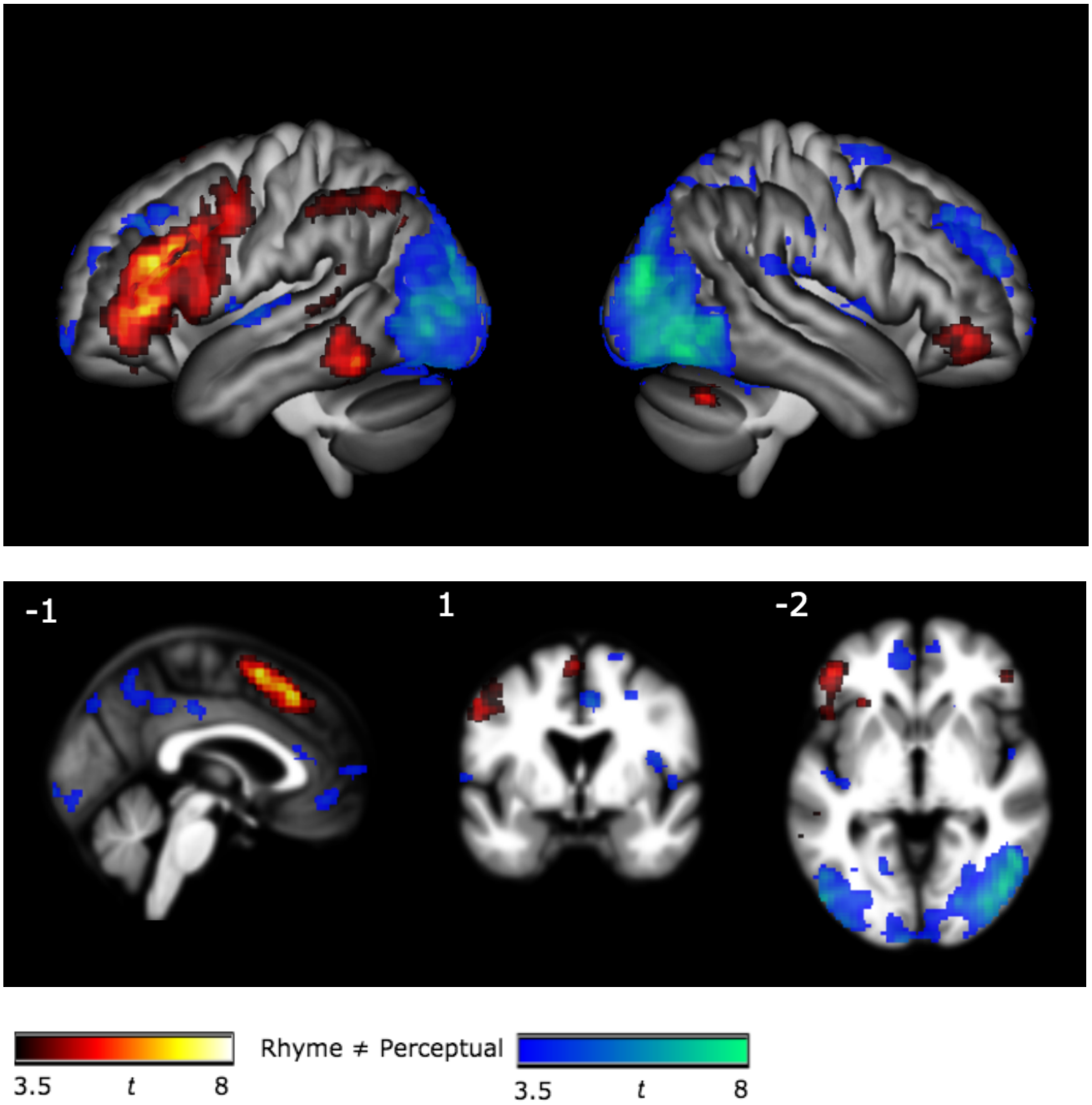
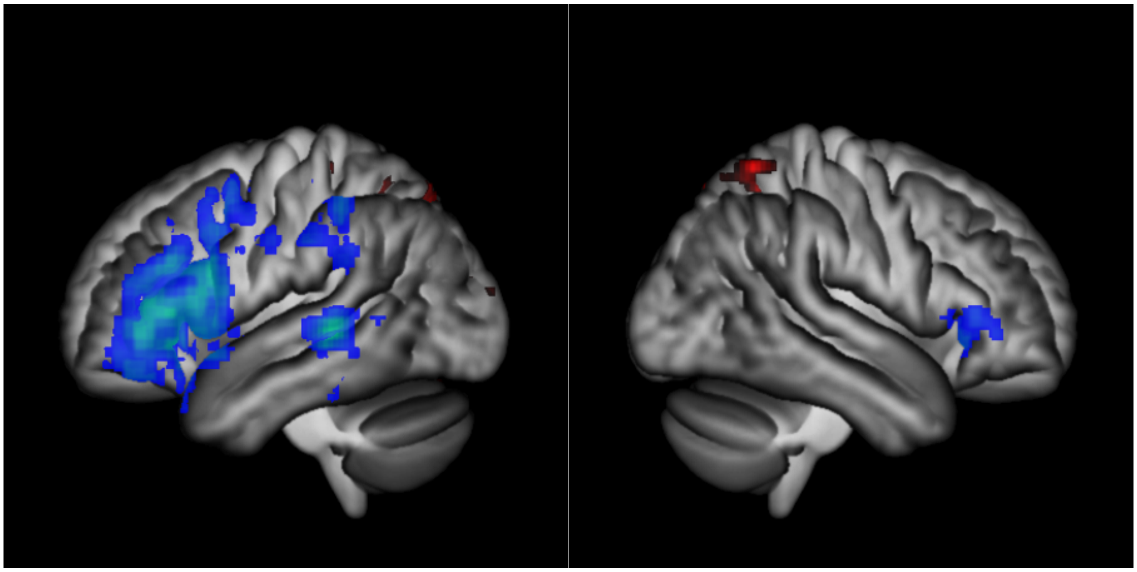


Figure 3.2 Group analysis for the rhyme vs. perceptual control contrast in 31 patients with aphasia. Again, both tails are shown: hot colors indicate greater activation for the rhyme condition, cool colors the control condition.

Direct contrasts

Comparing patients and controls directly, patients showed significantly less activation in the regions corresponding to components of the typical language network. This was consistent with our prediction that the patients as a group would show decreased activation in the frontal, temporal, and parietal language processing areas as a result of their lesions. The region where patients appeared to show greater activation was the bilateral precuneus. However, this was likely to be less deactivation for patients compared to controls, rather than more activation, because this posterior midline region was deactivated for both the patient and control groups individually (see Figures 3.1 and 3.2).



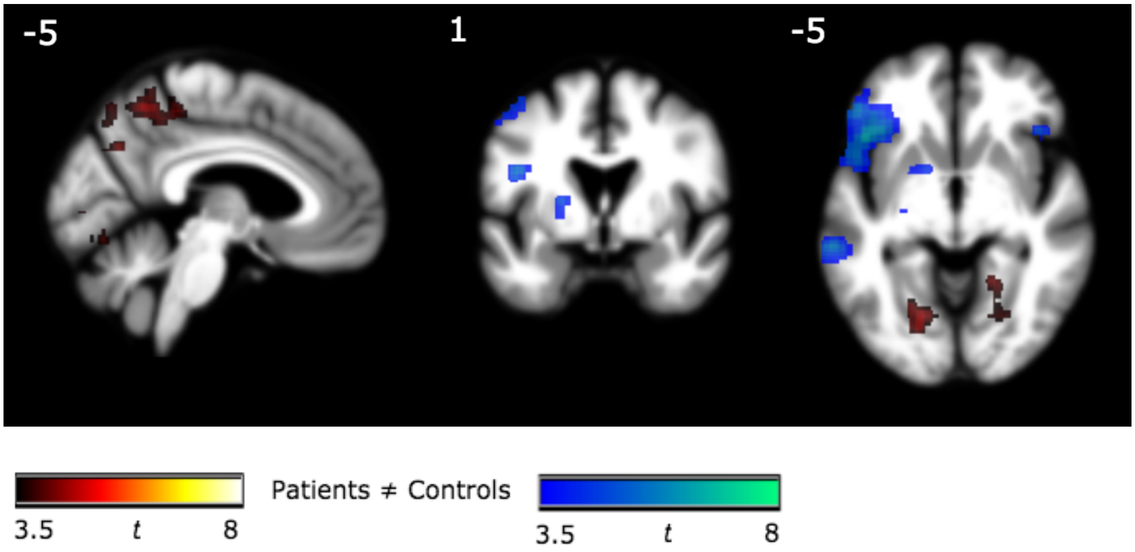
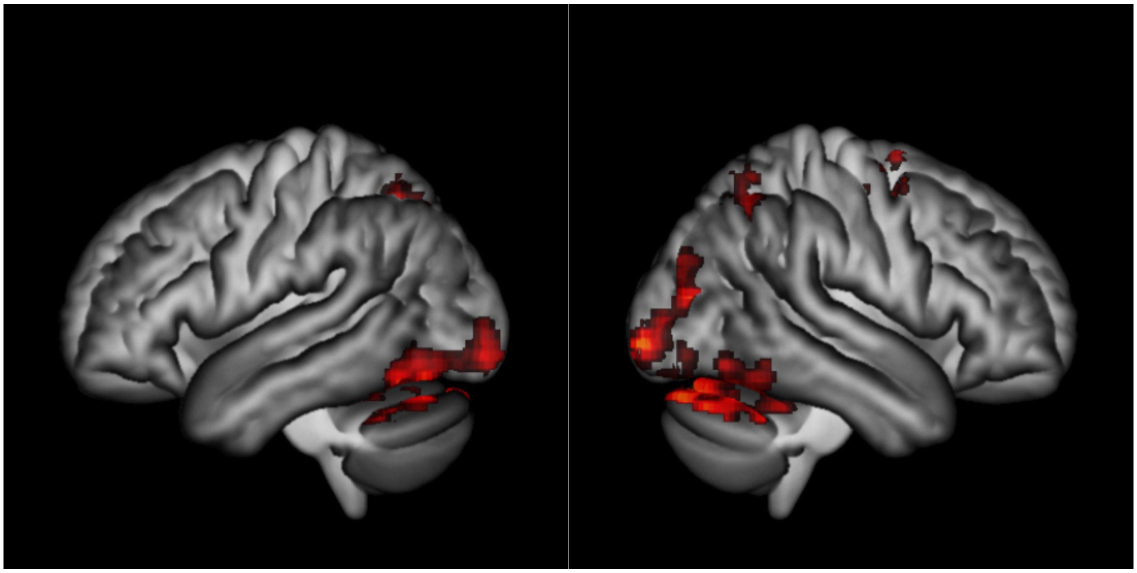


Figure 3.3 Between-groups contrast showing areas with greater activation for patients vs. controls. Hot colors indicate areas of greater activation for patients and cool colors indicate areas of greater activation for controls.

Covariate analyses

Activation maps including covariates with repetition scores included bilateral intraparietal sulcus, visual cortex, and superior frontal regions, a pattern that is consistent with the dorsal attention network.



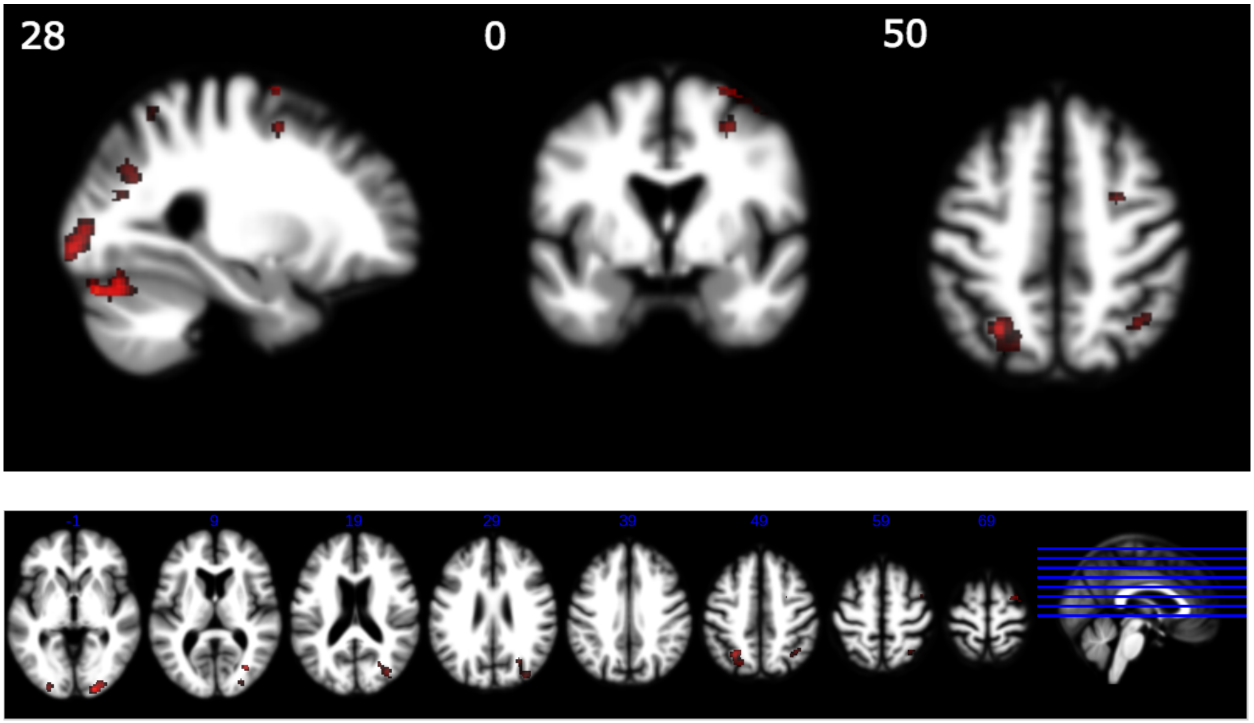


Figure 3.4 Areas correlated with better repetition scores.

This was made clear by overlaying the observed activations with a template of the brain's networks parcellated into seven systems (Yeo et al., 2011). The involvement of bilateral ventral visual areas indicated that patients who scored better on language assessment measures were applying more visual effort during the task. This suggests that patients who were more successful on the repetition subtest of the QAB were deploying more focus in scanning each pseudoword stimulus while building the phonological form.

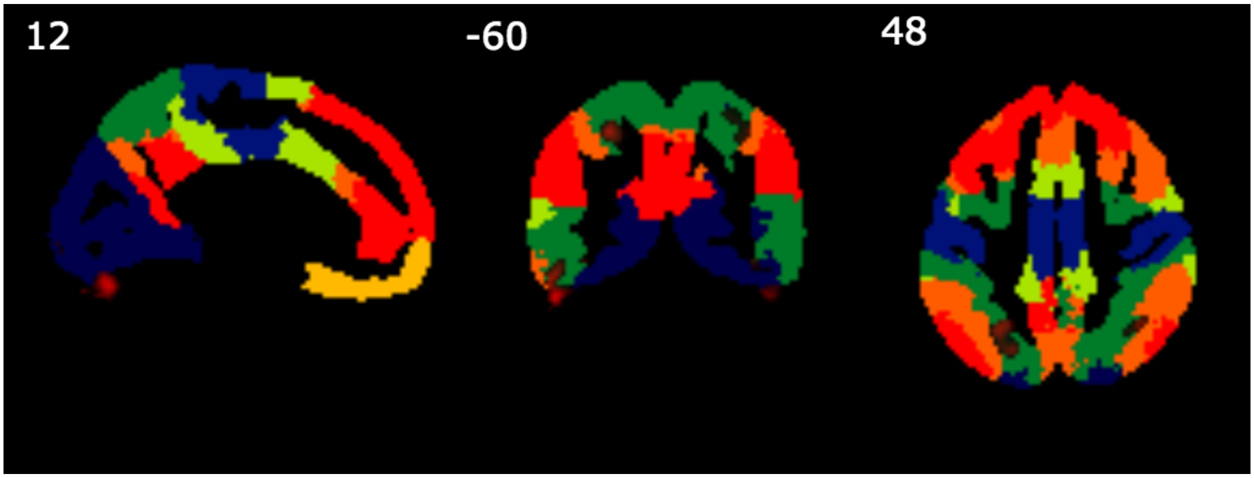
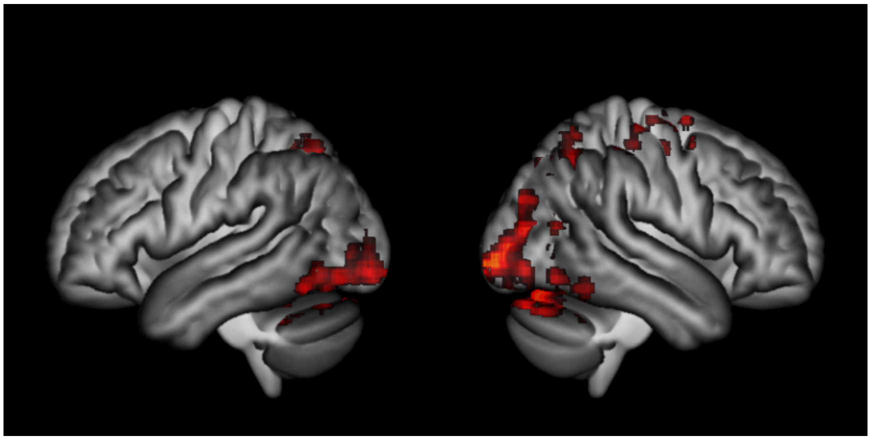


Figure 3.5 Repetition score correlation map overlaid on Yeo et al. (2011) template. Dark green sections of the template represent components of the dorsal attention network.

Consistent with the observation that repetition scores were strongly correlated with subjective ratings of each patient's phonological ability, the activation map correlating activity with perceptual rating appears very similar to that of the repetition map (Fig. 3.6).



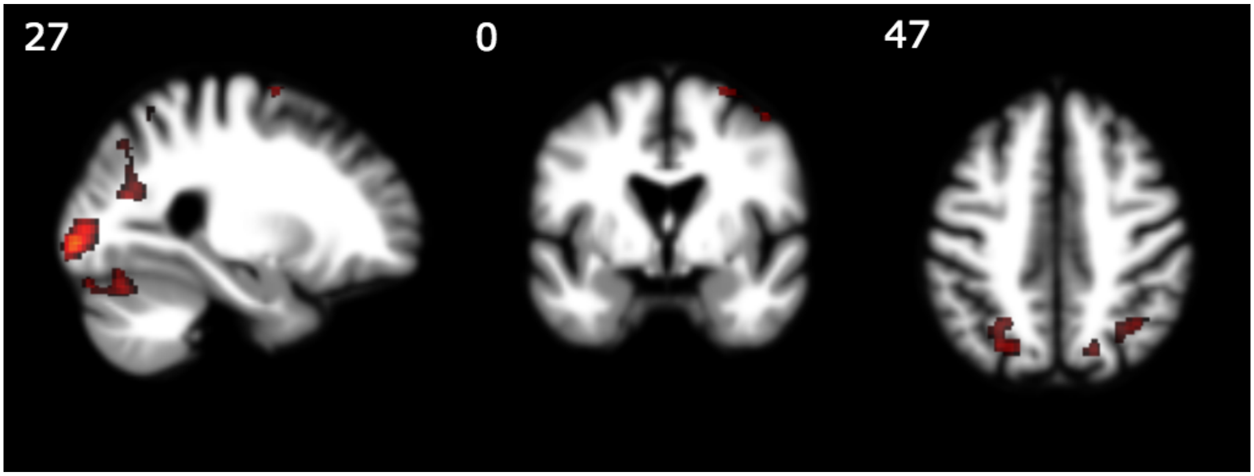


Figure 3.6 Areas correlated with subjective perceptual rating.

Correlating activation with accuracy on the rhyme task, it did not seem to be true that patients who performed better on the task showed patterns of activation that more closely approximated the typical rhyme network. In fact, there was only a small region of overlap between the typical network and the network that was more active in patients who did better in the rhyme task. The right pars triangularis was the single area that was more active for patients who achieved higher accuracy.

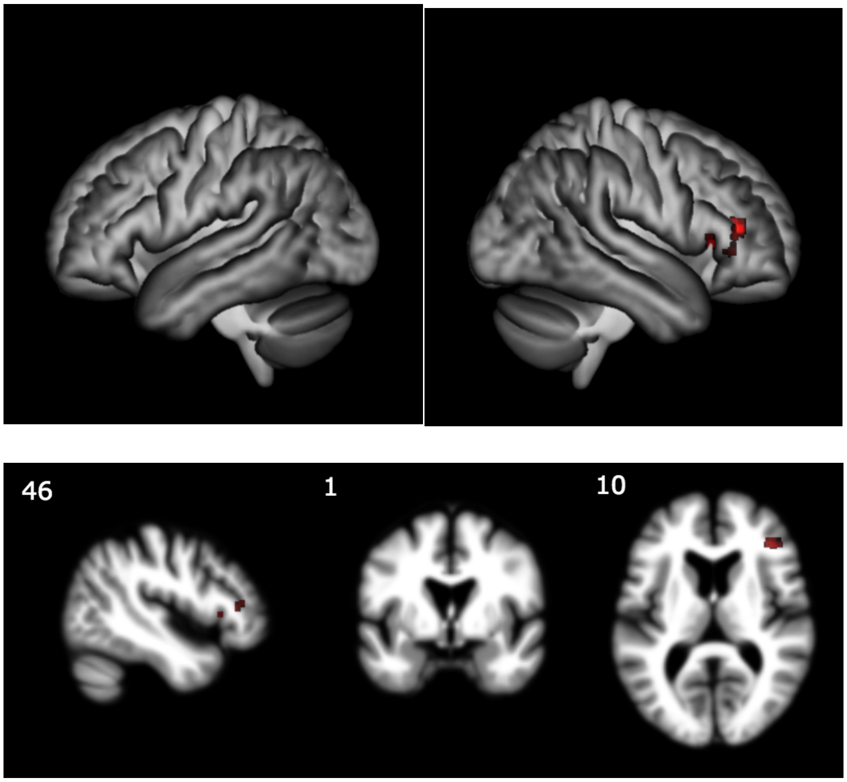


Figure 3.7 Areas correlated with greater rhyme task accuracy.

Activation in anterior frontal areas was correlated with the phonemic paraphasia score derived from the QAB. Interestingly, this activation map was dissimilar from that of the repetition scores. This suggested that, while repetition and phonemic paraphasias may both be related to phonological skill, they did not index it in identical ways. Contrary to our predictions, neither the repetition nor phonemic paraphasia scores from the QAB were correlated with increased activity in the supramarginal or precentral gyri.

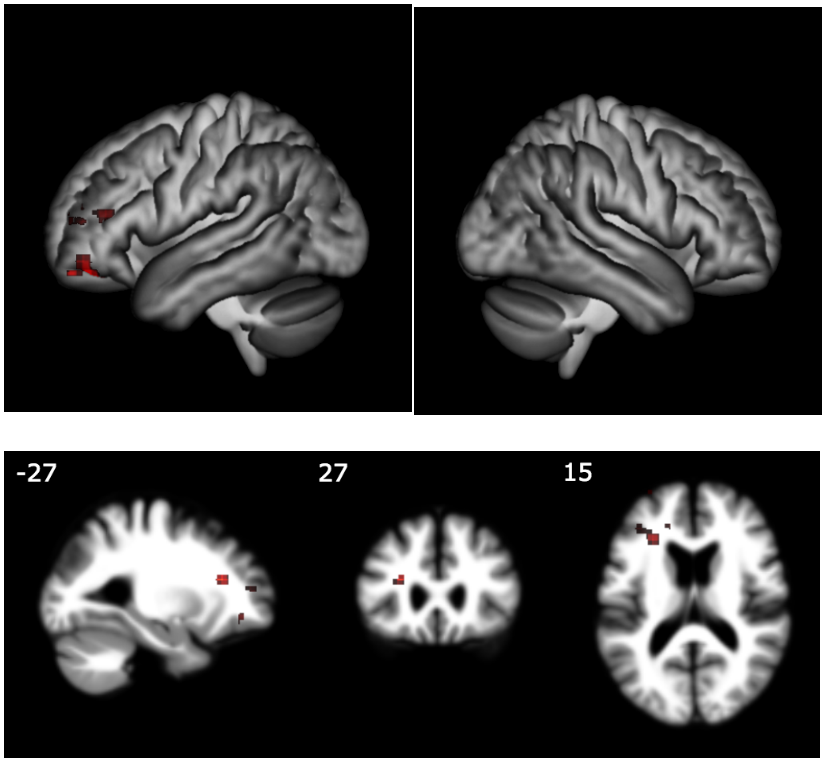


Figure 3.8 Areas correlated with phonemic paraphasia score.

We created a composite measure by combining the accuracy, repetition, and phonological paraphasia scores in order to better capture each patient's phonological ability. However, this measure did not demonstrate any results that were not captured by the measures individually.

A group analysis on patients who did and did not have damage to the typical phonological areas illustrated the regions that became more engaged when the frontoparietal areas typically associated with phonological function were disturbed. Patients with at least 20% of a region lesioned were considered to be damaged in that region. However, almost none of these comparisons survived $p < 0.005$ thresholding. The only activations that reached significance were modest clusters in the right occipital lobe and intraparietal sulcus for the negative tail of the precentral damage contrast. In other words, patients with more spared precentral gyri showed more activation in these regions than did patients with more damaged precentral gyri. With respect to other contrasts, several trends emerged that did not reach significance. Comparing patients with and without supramarginal gyrus damage, those with damage showed more activation in bilateral posterior superior temporal sulcus and anterior middle temporal sulcus. Comparing patients with and without precentral gyrus damage, those with damage showed more activation in bilataeral anterior frontal regions, bilateral insula, left postcentral gyrus, and left supramarginal gyrus.

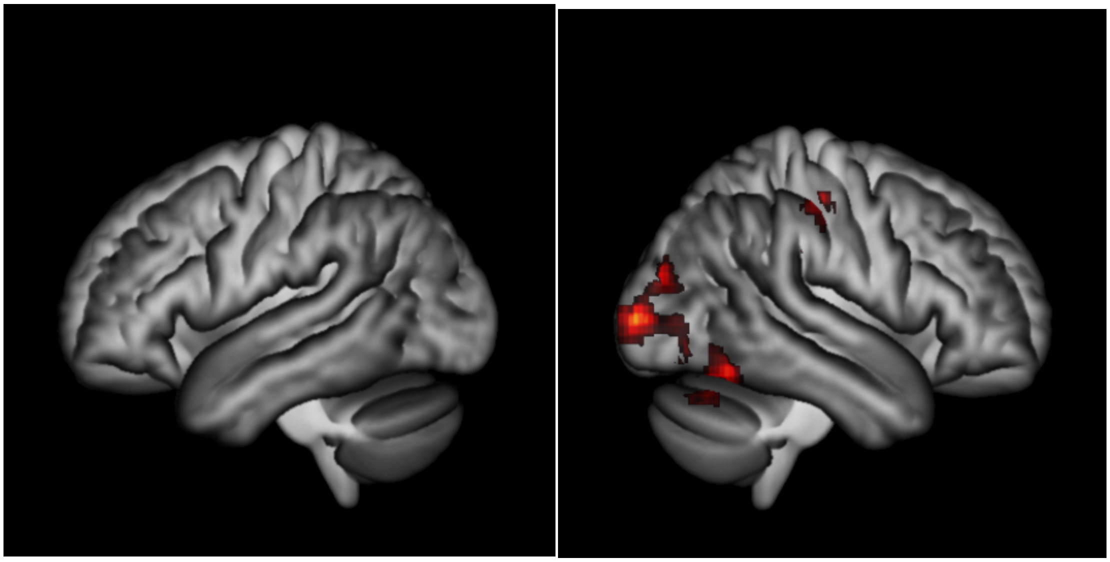


Figure 3.9 Contrast for patients with more spared precentral gyrus vs. patients with damaged precentral gyrus.

DISCUSSION

Overall, patients with aphasia performing language tasks show activation approximating the neurotypical language network. The rhyme paradigm was indeed successful in engaging the supramarginal gyrus and ventral precentral gyrus in this population. Repetition score appeared to be the most appropriate index of phonological encoding ability, as assessed by comparison to a direct perceptual characterization of patients' speech. Repetition performance is correlated with activity in the dorsal attention network for patients, suggesting that, given deficits in phonological processing capacities, the damaged brain applies a domain-general attention system to support operations on phonological stimuli.

Domain-general networks

The dorsal attention network consists of bilateral intraparietal sulcus and the frontal eye fields localized to the posterior superior frontal sulci (Corbetta et al., 2002; 2008). This network is thought to be involved in the voluntary direction of attention. It influences the processing of sensory stimuli based on previous experience, thereby enacting top-down modulation of sensory representations. This goal-directed attention network coexists with a ventral attention network thought to be responsible for mediation of behaviorally-relevant stimuli, such as pain.

Our findings indicate that patients with damage to phonological processing regions are more reliant on the dorsal attention network to perform well on language tasks. They may have to employ domain-general attentional processes that neurologically normal controls do not need to implement as strongly in order to succeed on these tasks. Where individuals without brain damage can quickly grasp linguistic stimuli, patients seem to rely more heavily on processes that are not specific to language in order to read, study, and act on the same stimuli.

Fedorenko and Thompson-Schill (2014) propose an organization of language composed of "core" and "periphery" systems, whereby core areas are specialized and co-activate only with each other during language processing, and peripheral areas co-activate with the core areas and other specialized systems as necessary. They argue for language as a specialized process that interacts in this way with executive systems. In the context of brain damage, then, a dysfunctional core system may prompt peripheral systems to be more active than usual in order to support the processing that a fully functional core network would be capable of accomplishing with less peripheral involvement.

Phonological processing with impaired frontoparietal regions

When patients were divided into damaged vs. undamaged groups with respect to their phonological processing regions, almost no activations reached significance after $p < 0.005$ thresholding. As such, while the activations we did observe may point to some potential patterns of compensatory function, these patterns may be fairly subtle and should be interpreted with caution. The sole significant clusters of activation were localized to the right intraparietal sulcus and occipital lobe for the spared vs. damaged precentral gyrus contrast. These clusters were nevertheless not very robust and may reflect activity other than genuine and stable patterns of functional reorganization.

Subjective and objective measures

The ratings produced through subjective impressions of the patients' language abilities were highly aligned with repetition scores (note that they were also extremely consistent with overall QAB score, a much more broad measure of language ability). This is evidence that repetition score is strongly based in the reality of the patient's deficit.

Interestingly, the behavioral measures did not all give similar results in terms of representing the patients' capabilities. Phonological paraphasia score was moderately correlated with perceptual rating and better performance was associated with increased activation in left anterior frontal regions. However, repetition score far outstripped this measure in alignment with subjective ratings, and the activation map produced with the phonological paraphasia covariate was quite dissimilar from the repetition and QAB covariate maps. This may be because phonological paraphasia score is based on a small speech sample from each patient. Moreover, phonological encoding deficits may occur in ways that do not manifest in paraphasia score. For example, having failed to successfully construct a word's phonological representation, the patient may choose a simpler word instead, or refrain from attempting to produce the word at all.

Accuracy on the rhyme task would seem to be a promising candidate for an appropriate indicator of phonological encoding ability. However, in practice, it was not strongly associated with other behavioral measures. Most critically, it showed a very limited relationship with perceptual rating of the patients' abilities. A credible proxy for phonological encoding should not be dissociated from a blinded, knowledgeable researcher's judgments of a patient's speech production. Repetition score may have succeeded where rhyme accuracy fell short because of the many variegated factors that contribute to in-scanner performance: high accuracy depends on not only phonological processing skill, but also reading skill, task-related executive skills, and tolerance of scanner conditions, among others.

Another interesting finding was that the composite measure built from the repetition, accuracy, and phonological paraphasia measures together was not successful in informing the results beyond what each measure illustrated individually. This demonstrates that a multidimensional measure is not necessarily superior to simpler ones, perhaps because the component measures varied in their ability to represent the relevant behavior.

CHAPTER IV

General discussion and conclusion

EFFECTIVENESS OF ADAPTIVE LANGUAGE MAPPING

The paradigms we developed outperform existing language mapping paradigms in reliability and sensitivity to language processing regions. Moreover, they are successful even at the individual level and load on specific domains of language such that the resulting activation maps go beyond indexing undifferentiated “eloquent cortex”. Much of the literature regarding presurgical mapping does not distinguish between different aspects of the language network. Presurgical procedures tend to approach “eloquent cortex” as a homogenous portion of cortex that produces speech arrest when stimulated (Gallen et al., 1995; McDonald et al., 1999; Danks, Aglio, Gugino, & Black, 2000). Our tasks allow for more targeted localization of language areas while maintaining high standards of reliability and validity. In addition to these psychometric advantages, the adaptive design both controls for difficulty effects and makes the paradigms viable for language-impaired populations. These qualities make our paradigms a major step forward in the contexts of presurgical mapping and language-related research.

Applying the paradigms to individuals with aphasia demonstrated that language function in the damaged brain is not restricted to approximation of the undamaged network. The increased activation in the dorsal attention network correlated with better repetition score suggests that domain-general networks are recruited more strongly when the typical language network is impaired. This study constitutes the first demonstration of robust evidence that systems nonspecific to language support language function in patients with aphasia. Several studies have previously reported recruitment of domain-general networks during recovery from post-stroke aphasia (DeMarco, Wilson, Rising, Rapcsak, & Beeson, 2018; Brownsett et al., 2014; Geranmayeh, Chau, Wise, Leech, & Hampshire, 2017), but confidence in these findings is undermined by weaknesses in the study or analysis design. DeMarco et al. (2018) illustrates activity in regions involved in effort and attention, but these results were only demonstrated in a single patient. While the single subject design does not invalidate the findings, it inherently limits the generalizability of the conclusions. The contrasts in Brownsett et al. (2014) are interesting in their attempt to control for performance using noise vocoded speech, but the language contrast in patients seems confounded by motor activity. Geranmayeh et al. (2017) describe activation in the presupplementary motor area/dorsal anterior cingulate cortex correlated with speech production recovery. They interpret this preSMA/dACC activity as domain-general contributions to task processing. However, given that the authors did not observe activation in other parts of the network, it is unclear whether or not their ROI accurately represented the preSMA/dACC component of the cingulo-opercular network. It may, for example, be capturing the medial frontal component of the semantic network instead. As such, the current study provides unprecedented evidence of domain-general support of language-specific processing, and moreover, offers improved tools for clinical applications and further research in the context of language network localization.

LIMITATIONS

Several limitations remain in our study. First, despite the wide range of situations that the paradigms are applicable to, they are still inappropriate in certain contexts. Patients whose deficits are severe enough will still have difficulty learning and performing the tasks. Four of our five excluded patients were unable to viably complete the paradigms for this reason. One struggled with the stimuli due to questionable baseline reading ability. Adaptive Language Mapping features some auditory options to mitigate instances of exclusion based on visual or reading problems, but further development of alternate versions of the tasks (including translating the stimuli into other languages) will help maximize the potential of their use.

Second, as described in Chapter Three, the study in its current form cannot unambiguously disentangle phonological encoding from related phonological processes. It is likely that at least some of the observed activation reflects the orthographic decoding and metalinguistic operations that were involved in the tasks. While this means that interpretation of the findings cannot be as straightforward as a one-to-one mapping between a brain region and the process of phonological encoding in isolation, our results still undeniably illustrate a relationship between phonological processes and particular areas of cortex. It remains true that the phonological operations of phoneme sequencing and syllabification are required to perform the tasks, and these operations must be reflected in the activation patterns we observed.

Third, the finding that repetition score was strongly correlated with overall QAB score somewhat undermines the idea that it was successful in capturing phonological ability specifically. It may be that the repetition subtest score is more reflective of overall aphasia severity than we thought.

FUTURE WORK

Future directions include refining the paradigms and applying them to other aspects of language research. As previously mentioned, options for other modalities and languages can be further refined and developed. Further work maximizing the efficacy of the tasks can also empirically characterize each difficulty level and assess their ability to successfully load on greater processing skill as level increases. A study unpacking the effects of hard compared to easy levels would elucidate outstanding questions about how the brain deploys effort, and whether or not this effort is represented differently depending on level or function.

The paradigms can also be applied to patients with aphasia at different points in their recovery, with the goal of comparing groups from various time points to each other. This would produce insights into how functional localization changes during the trajectory of recovery. Establishing the nature of this pattern of change would not only be of interest to researchers, but also have implications for individuals with aphasia receiving speech-language therapy. A firmer understanding of which brain regions are most involved in each stage of recovery would help inform what type of intervention might synergize with natural post-damage changes to most effectively improve patient outcomes.

Patients who require brain surgery can also perform the tasks as part of their presurgical mapping procedures, and the postsurgical outcomes can be compared to those of patients who had performed other language mapping tasks. While a study of this kind would be time-consuming and complex, due to the ethical logistics of having different groups undergoing different presurgical mapping methods, it is not implausible and would greatly improve the field's understanding of which techniques actually lead to better patient outcomes. These are all important next steps for improving and validating the Adaptive Language Mapping paradigms as the state of the art for identifying the language network.

REFERENCES

Ashburner, J., & Friston, K. J. (2005). Unified segmentation. *NeuroImage, 26,* 839–851.

Awad, M., Warren, J. E., Scott, S. K., Turkheimer, F. E., & Wise, R. J. (2007). A common system for the comprehension and production of narrative speech. *Journal of Neuroscience*, *27*, 11455-11464.

Baciu, M., Kahane, P., Minotti, L., Charnallet, A., David, D., Le Bas, J. F., & Segebarth, C. (2001). Functional MRI assessment of the hemispheric predominance for language in epileptic patients using a simple rhyme detection task. *Epileptic Disorders, 3,* 117–124.

Baciu, M. V., Watson, J. M., Maccotta, L., McDermott, K. B., Buckner, R. L., Gilliam, F. G., & Ojemann, J. G. (2005). Evaluating functional MRI procedures for assessing hemispheric language dominance in neurosurgical patients. *Neuroradiology, 47,* 835–844.

Baldo, J. V., & Dronkers, N. F. (2006). The role of inferior parietal and inferior frontal cortex in working memory. *Neuropsychology*, *20*, 529.

Beckmann, C. F., & Smith, S. M. (2004). Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE Transactions on Medical Imaging, 23,* 137–152.

Belin, P., Zilbovicius, M., Remy, P., Francois, C., Guillaume, S., Chain, F., Rancurel, G., & Samson, Y. (1996). Recovery from nonfluent aphasia after melodic intonation therapy A PET study. *Neurology*, *47*, 1504-1511.

Berl, M. M., Mayo, J., Parks, E. N., Rosenberger, L. R., VanMeter, J., Ratner, N. B., Vaidya, C.J., & Gaillard, W. D. (2014). Regional differences in the developmental trajectory of lateralization of the language network. *Human brain mapping*, *35*, 270-284.

Billingsley, R. L., McAndrews, M. P., Crawley, A. P., & Mikulis, D. J. (2001). Functional MRI of phonological and semantic processing in temporal lobe epilepsy. *Brain, 124,* 1218–1227.

Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S. F., Rao, S. M., & Cox, R. W. (1999). Conceptual processing during the conscious resting state: A functional MRI study. *Journal of Cognitive Neuroscience*, *11*, 80–93.

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. *Nature Neuroscience*, 7, 295–301.

Binder, J. R., Medler, D. A., Desai, R., Conant, L. L., & Liebenthal, E. (2005). Some neurophysiological constraints on models of word naming. *NeuroImage*, 27, 677–693.

Binder, J. R., Swanson, S. J., Hammeke, T. A., & Sabsevitz, D. S. (2008). A comparison of five fMRI protocols for mapping speech comprehension systems. *Epilepsia, 49,* 1980–1997.

Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex*, *19*, 2767-2796.

Booth, J. R., Burman, D. D., Meyer, J. R., Gitelman, D. R., Parrish, T. B., & Mesulam, M. M. (2002). Functional anatomy of intra- and cross-modal lexical tasks. *NeuroImage, 16,* 7–22.

Booth, J. R., Lu, D., Burman, D. D., Chou, T.-L., Jin, Z., Peng, D.-L., Zhang, L., Ding, G.-S., Deng, Y., & Liu, L. (2006). Specialization of phonological and semantic processing in Chinese word reading. *Brain Research, 1071,* 197–207.

Bradshaw, A. R., Thompson, P. A., Wilson, A. C., Bishop, D. V. M., & Woodhead, Z. V. J. (2017). Measuring language lateralisation with different language tasks: a systematic review. *PeerJ, 5,* e3929.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10,* 433–436.

Braun, A. R., Varga, M., Stager, S., Schulz, G., Selbie, S., Maisog, J. M., … Ludlow, C. L. (1997). Altered patterns of cerebral activity during speech and language production in developmental stuttering: an $H_2{}^{15}O$ positron emission tomography study. *Brain, 120,* 761–784.

Braver, T. S., Cohen, J. D., Nystrom, L. E., Jonides, J., Smith, E. E., & Noll, D. C. (1997). A parametric study of prefrontal cortex involvement in human working memory. *Neuroimage*, *5*, 49-62.

Burton, M. W., LoCasto, P. C., Krebs-Noble, D., & Gullapalli, R. P. (2005). A systematic investigation of the functional neuroanatomy of auditory and visual phonological processing. *NeuroImage, 26,* 647–661.

Broca, P. (1861). Remarks on the seat of the faculty of articulated language, following an observation of aphemia (loss of speech). *Bulletin de la Société Anatomique*, *6*, 330-57.

Brownsett, S. L. E., Warren, J. E., Geranmayeh, F., Woodhead, Z., Leech, R., & Wise, R. J. S. (2014). Cognitive control and its impact on recovery from aphasic stroke. *Brain, 137,* 242–254.

Caramazza, A., & Zurif, E. B. (1976). Dissociation of algorithmic and heuristic processes in language comprehension: Evidence from aphasia. *Brain and Language*, *3*, 572-582.

Casilio, M., Rising, K., Beeson, P.M., Bunton, K., Wilson, S.M. Auditory-perceptual rating of connected speech in aphasia. *Am J Speech Lang Pathol* 2019; in press.

Clements, A. M., Rimrodt, S. L., Abel, J. R., Blankner, J. G., Mostofsky, S. H., Pekar, J. J., Denckla, M.B., & Cutting, L. E. (2006). Sex differences in cerebral laterality of language and visuospatial processing. *Brain and Language, 98,* 150–158.

Coltheart, M. (1981): The MRC psycholinguistic database. *Q J Exp Psychol A,* 33A:497–505.

Corbetta, M., Kincade, J. M., & Shulman, G. L. (2002). Neural Systems for Visual Orienting and Their Relationships to Spatial Working Memory. *Journal of Cognitive Neuroscience*, *14*(3), 508–523.

Corbetta, M., Patel, G., & Shulman, G. L. (2008). The Reorienting System of the Human Brain: From Environment to Theory of Mind. *Neuron*, *58*(3), 306–324.

Corbetta, M., Ramsey, L., Callejas, A., Baldassarre, A., Hacker, C. D., Siegel, J. S., & Connor, L. T. (2015). Common behavioral clusters and subcortical anatomy in stroke. *Neuron*, *85*, 927-941.

Cousin, E., Peyrin, C., Pichat, C., Lamalle, L., Le Bas, J.-F., & Baciu, M. (2007). Functional MRI approach for assessing hemispheric predominance of regions activated by a phonological and a semantic task. *European Journal of Radiology, 63,* 274–285.

Cox, R. W. (1996). AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research, 29,* 162–173.

Crinion, J. T., Lambon-Ralph, M. A., Warburton, E. A., Howard, D., & Wise, R. J. (2003). Temporal lobe regions engaged during normal speech comprehension. *Brain*, *126*, 1193-1201.

Crinion, J., Price, C.J. (2005). Right anterior superior temporal activation predicts auditory sentence comprehension following aphasic stroke. *Brain*, *128*, 2858–2871.

Danks, R.A., Aglio, L.S., Gugino, L.D., & Black, P.M. (2000). Craniotomy under local anesthesia and monitored conscious sedation for the resection of tumors involving eloquent cortex. *Journal of Neuro-Oncology*, *49*, 131-139.

Dax, G. (1863). Observations tendant à prouver la coïncidence constante des dérangements de la parole avec une lésion de l'hémisphère gauche du cerveau. *CR Acad Sci Hebd Seances Acad Sci*, *61*, 534.

DeMarco, A. T. (2016). *Neural substrates of phonological processing in chronic aphasia from stroke* (Unpublished doctoral dissertation). University of Arizona, Tucson, USA.

DeMarco, A. T., Wilson, S. M., Rising, K., Rapcsak, S. Z., & Beeson, P. M. (2018). The neural substrates of improved phonological processing following successful treatment in a case of phonological alexia and agraphia. *NeuroCase, 24,* 31–40.

Farah, M. J., Gazzaniga, M. S., Holtzman, J. D., & Kosslyn, S. M. (1985). A left hemisphere basis for visual mental imagery? *Neuropsychologia, 23,* 115–118.

Fedorenko, E., Hsieh, P.-J., Nieto-Castañón, A., Whitfield-Gabrieli, S., & Kanwisher, N. (2010). New method for fMRI investigations of language: Defining ROIs functionally in individual subjects. *Journal of Neurophysiology, 104,* 1177–1194.

Fedorenko, E., Duncan, J., & Kanwisher, N. (2013). Broad domain generality in focal regions of frontal and parietal cortex. *Proceedings of the National Academy of Sciences of the United States of America, 110,* 16616–16621.

Fedorenko, E., Duncan, J., & Kanwisher, N. (2012). Language-selective and domain-general regions lie side by side within Broca's area. *Current Biology*, 22, 2059–2062.

Fiez, J. A., Balota, D. A., Raichle, M. E., & Petersen, S. E. (1999). Effects of lexicality, frequency, and spelling-to-sound consistency on the functional anatomy of reading. *Neuron*, *24*, 205-218.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state." *Journal of Psychiatric Research, 12,* 189–198.

Forkel, S. J., Thiebaut de Schotten, M., Dell'Acqua, F., Kalra, L., Murphy, D. G., Williams, S. C., & Catani, M. (2014). Anatomical predictors of aphasia recovery: a tractography study of bilateral perisylvian language networks. *Brain*, *137*, 2027-2039.

Fridriksson, J., Baker, J. M., & Moser, D. (2009). Cortical mapping of naming errors in aphasia. *Human brain mapping*, *30*, 2487-2498.

Fridriksson, J., Kjartansson, O., Morgan, P. S., Hjaltason, H., Magnusdottir, S., Bonilha, L., & Rorden, C. (2010). Impaired speech repetition and left parietal lobe damage. *Journal of Neuroscience, 30,* 11057–11061.

Fridriksson, J., Richardson, J.D., Fillmore, P., Cai, B. (2012): Left hemisphere plasticity and aphasia recovery. *NeuroImage*, *60*, 854–863.

Fridriksson, J., Yourganov, G., Bonilha, L., Basilakos, A., Den Ouden, D.-B., & Rorden, C. (2016). Revealing the dual streams of speech processing. *Proceedings of the National Academy of Sciences of the United States of America, 113,* 15108–15113.

Gaillard, William D., L. Balsamo, B. Xu, C. McKinney, P. H. Papero, S. Weinstein, J. Conry, P. L., Pearl, B. Sachs, S. Sato, L. G. Venzina, C. Frattali, & W. H. Theodore. (2004). fMRI language task panel improves determination of language dominance. *Neurology,* 63, 1403–1408.

Gaillard, W. D., Berl, M. M., Moore, E. N., Ritzl, E. K., Rosenberger, L. R., Weinstein, S. L., Conry, J.A., Pearl, P.L., Ritter, F.F., Sato, S., Vezina, L.G., Vaidya, C.J., Wiggs, E., Fratalli, C., Risse, G., Ratner, N.B., Gioia, G., & Theodore, W. H. (2007). Atypical language in lesional and nonlesional complex partial epilepsy. *Neurology, 69,* 1761–1771.

Gallen, C.C., Schwartz, B.J., Bucholz, R.D., Malik, G., Barkley, G.L., Smith, J., Tung, H., Copeland, B., Bruno, L., Assam, S., Hirschkoff, E., & Bloom, F. (1995). Presurgical localization of functional cortex using magnetic source imaging. *Journal of Neurosurgery, 82*, 988-994.

García-Pérez, M. A. (1998). Forced-choice staircases with fixed step sizes: asymptotic and small-sample properties. *Vision research*, *38*, 1861-1881.

Gazzaniga, M. S. (1995). Consciousness and the cerebral hemispheres. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 1391–1400). Cambridge, MA: MIT Press.

Geranmayeh, F., Brownsett, S. L. E., & Wise, R. J. S. (2014). Task-induced brain activity in aphasic stroke patients: What is driving recovery? *Brain, 137,* 2632–2648.

Geranmayeh, F., Chau, T. W., Wise, R. J. S., Leech, R., & Hampshire, A. (2017). Domain-general subregions of the medial prefrontal cortex contribute to recovery of language after stroke. *Brain, 140,* 1947–1958.

Geva, S., Jones, P. S., Crinion, J. T., Price, C. J., Baron, J.-C., & Warburton, E. A. (2011). The neural correlates of inner speech defined by voxel-based lesion-symptom mapping. *Brain, 134,* 3071–3082.

Gitelman, D. R., Nobre, A. C., Sonty, S., Parrish, T. B., & Mesulam, M.-M. (2005). Language network specializations: An analysis with parallel task designs and functional magnetic resonance imaging. *NeuroImage, 26,* 975–985.

Gough, P. M., Nobre, A. C., & Devlin, J. T. (2005). Dissociating linguistic processes in the left inferior frontal cortex with transcranial magnetic stimulation. *Journal of Neuroscience, 25,* 8010–8016.

Grabski, K., Tremblay, P., Gracco, V. L., Girin, L., & Sato, M. (2013). A mediating role of the auditory dorsal pathway in selective adaptation to speech: A state-dependent transcranial magnetic stimulation study. *Brain Research, 1515,* 55–65.

Gross, W. L., & Binder, J. R. (2014). Alternative thresholding methods for fMRI data optimized for surgical planning. *NeuroImage, 84,* 554–561.

Harrington, G. S., Buonocore, M. H., & Farias, S. T. (2006). Intrasubject reproducibility of functional MR imaging activation in language tasks. *American Journal of Neuroradiology*, *27*, 938-944.

Hartwigsen, G., & Saur, D. (2017). Neuroimaging of stroke recovery from aphasia – Insights into plasticity of the human language network. *NeuroImage*. In press.

Heiss, W. D., Kessler, J., Karbe, H., Fink, G. R., & Pawlik, G. (1993). Cerebral glucose metabolism as a predictor of recovery from aphasia in ischemic stroke. *Archives of Neurology*, *50*, 958-964.

Heiss, W.D., Karbe, H., Weber-Luxenburger, G., Herholz, K., Kessler, J., Pietrzyk, U., Pawlik, G. (1997): Speech-induced cerebral metabolic activation reflects recovery from aphasia. *J Neurol Sci*, *145*, 213–217.

Heiss, W.D., Kessler, J., Thiel, A., Ghaemi, M., Karbe, H. (1999): Differential capacity of left and right hemispheric areas for compensation of poststroke aphasia. *Ann Neurol*, *45*, 430–438.

Heiss, W.D., & Thiel, A. (2006). A proposed regional hierarchy in recovery of post stroke aphasia. *Brain and Language*, *98*, 118–123.

Hope, T. M., Leff, A. P., Prejawa, S., Bruce, R., Haigh, Z., Lim, L., Ramsden, S., Oberhuber, M., Ludersdorfer, P., Crinion, J., Seghier, M. L., & Price, C.J. (2017).

Right hemisphere structural adaptation and changing language skills years after left hemisphere stroke. *Brain*, *140*, 1718-1728.

Howard, D. & Patterson, K. Pyramids and palm trees: A test of semantic access from pictures and words. Bury St. Edmunds: Thames Valley; 1992.

Ito, S., Stuphorn, V., Brown, J. W., & Schall, J. D. (2003). Performance monitoring by the anterior cingulate cortex during saccade countermanding. *Science, 302,* 120–122.

Jobard, G., Crivello, F., & Tzourio-Mazoyer, N. (2003). Evaluation of the dual route theory of reading: A metanalysis of 35 neuroimaging studies. *NeuroImage, 20,* 693–712.

Jonides, J., Schumacher, E. H., Smith, E. E., Koeppe, R. A., Awh, E., Reuter-Lorenz, P. A., Marshuetz, C., & Willis, C. R. (1998). The role of parietal cortex in verbal working memory. *Journal of neuroscience*, *18*, 5026-5034.

Kelly, R. E., Alexopoulos, G. S., Wang, Z., Gunning, F. M., Murphy, C. F., Morimoto, S. S., Kanellopoulos, D., Jia, Z., Lim, K.O., & Hoptman, M. J. (2010). Visual inspection of independent components: Defining a procedure for artifact removal from fMRI data. *Journal of Neuroscience Methods, 189,* 233–245.

Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior Research Methods, 42,* 627–633.

Kim, S. H., & Jang, S. H. (2013). Prediction of aphasia outcome using diffusion tensor tractography for arcuate fasciculus in stroke. *American Journal of Neuroradiology*, *34*, 785-790.

Kiran, S., Ansaldo, A., Bastiaanse, R., Cherney, L.R., Howard, D., Faroqi-Shah, Y., Meinzer, M., Thompson, C.K. (2013): Neuroimaging in aphasia treatment research: standards for establishing the effects of treatment. *NeuroImage*, *76*, 428–435.

Knecht, S., Jansen, A., Frank, A., van Randenborgh, J., Sommer, J., Kanowski, M., & Heinze, H. J. (2003). How atypical is atypical language dominance? *NeuroImage, 18,* 917–927.

Koechlin, E., & Jubault, T. (2006). Broca's area and the hierarchical organization of human behavior. *Neuron*, *50*, 963-974.

Kosslyn, S. M. (1988). Aspects of a cognitive neuroscience of mental imagery. *Science, 240,* 1621–1626.

Kümmerer, D., Hartwigsen, G., Kellmeyer, P., Glauche, V., Mader, I., Klöppel, S., Suchan, J., Karnath, H-O., Weiller, C., & Saur, D. (2013). Damage to ventral and dorsal language pathways in acute aphasia. *Brain*, *136*, 619-629.

Kuperman, V., Stadthagen-Gonzalez, H., Brysbaert, M. (2012): Age-of-acquisition ratings for 30,000 English words. *Behav Res Methods*, *44*, 978–990.

Laska, A. C., Hellblom, A., Murray, V., Kahan, T., & Von Arbin, M. (2001). Aphasia in acute stroke and relation to outcome. *Journal of Internal Medicine, 249,* 413–422.

Leff, A., Crinion, J., Scott, S., Turkheimer, F., Howard, D., & Wise, R. (2002). A physiological change in the homotopic cortex following left posterior temporal lobe infarction. *Annals of neurology*, *51*, 553-558.

Levelt, W. J. M. (1989). *Speaking: from intention to articulation*. Cambridge, MA: MIT Press.

Levelt, W. J. M. (2001). Spoken word production: a theory of lexical access. *Proceedings of the National Academy of Sciences of the United States of America, 98,* 13464–13471.

Lurito, J. T., Kareken, D. A., Lowe, M. J., Chen, S. H. A., & Mathews, V. P. (2000). Comparison of rhyming and word generation with FMRI. *Human Brain Mapping, 10,* 99–106.

Mandera, P., Keuleers, E., & Brysbaert, M. (2017): Explaining human performance in psycholinguistic tasks with models of semantic similarity based on prediction and counting: A review and empirical validation. *J Mem Lang*, *92*, 57–78.

Martin, P. I., Naeser, M. A., Theoret, H., Tormos, J. M., Nicholas, M., Kurland, J., Fregni, F., Seekins, H., Doron, K., & Pascual-Leone, A. (2004, May). Transcranial magnetic stimulation as a complementary treatment for aphasia. *Seminars in speech and language, 25*,181-191.

McDermott, K. B., Petersen, S. E., Watson, J. M., & Ojemann, J. G. (2003). A procedure for identifying regions preferentially activated by attention to semantic and phonological relations using functional magnetic resonance imaging. *Neuropsychologia, 41,* 293–303.

McDonald, J.D., Chong, B.W., Lewine, J.D., Jones, G., Burr, R.B., McDonald, P.R., Koehler, S.B., Tsuruda, J., Orrison, W.W., & Heilbrun, M.P. (1999). Integration of preoperative and intraoperative functional brain mapping in a frameless stereotactic environment for lesions near eloquent cortex: Technical note. *Journal of Neurosurgery*, *90*, 591-598.

Mechelli, A., Gorno-Tempini, M. L., & Price, C. J. (2003). Neuroimaging studies of word and pseudoword reading: consistencies, inconsistencies, and limitations. *Journal of Cognitive Neuroscience, 15,* 260–271.

Meinzer, M., Beeson, P.M., Cappa, S., Crinion, J., Kiran, S., Saur, D., Parrish, T., Crosson, B., Thompson, C.K., Neuroimaging in Aphasia Treatment Research Workshop (2013): Neuroimaging in aphasia treatment research: Consensus and practical guidelines for data analysis. *NeuroImage*, *73*, 215–224.

Mirman, D., Chen, Q., Zhang, Y., Wang, Z., Faseyitan, O. K., Coslett, H. B., & Schwartz, M. F. (2015). Neural organization of spoken language revealed by lesion–symptom mapping. *Nature Communications, 6,* 6762.

Mohr, J.P. Broca's area and Broca's aphasia. In: Whitaker H, Whitaker H, editors. *Studies in neurolinguistics*. New York: Academic Press; 1976. p. 201–33.

Morrison, M. A., Churchill, N. W., Cusimano, M. D., Schweizer, T. A., Das, S., & Graham, S. J. (2016). Reliability of task-based fMRI for preoperative planning: A test-retest study in brain tumor patients and healthy controls. *PLoS One, 11,* e0149547.

Musso, M., Weiller, C., Kiebel, S., Müller, S. P., Bülau, P., & Rijntjes, M. (1999). Training-induced brain plasticity in aphasia. *Brain*, *122*, 1781-1790.

Nadeau, S.E. (2014): Neuroplastic mechanisms of language recovery after stroke. In: Tracy, JI, Hampstead, BM, Sathian, K, editors. Cognitive plasticity in neurologic disorders. Oxford: Oxford University Press. pp 61–84.

Naeser, M. A., Martin, P. I., Nicholas, M., Baker, E. H., Seekins, H., Kobayashi, M., Theoret, H., Fregni, F., Maria-Tormos, J., Kurland, J., Doron, K.W., & Pascual Leone, A. (2005). Improved picture naming in chronic aphasia after TMS to part of right Broca's area: an open-protocol study. *Brain and language*, *93*, 95-105.

Nelson, D., McEvoy, C., & Schreiber, T. (1998): The University of South Florida word association, rhyme, and word fragment norms. http://w3.usf.edu/FreeAssociation.

Newport, E.L., Landau, B., Seydell-Greenwald, A., Turkeltaub, P.E., Chambers, C.E., Dromerick, A.W., Carpenter, J., Berl, M.M., & Gaillard, W.D. (2017). Revisiting Lenneberg's hypotheses about early developmental plasticity: language organization after left-hemisphere perinatal stroke. *Biolinguistics*.

Nichols, T. E., & Holmes, A. P. (2002). Nonparametric permutation tests for functional neuroimaging: A primer with examples. *Human Brain Mapping, 15,* 1–25.

Oberhuber, M., Hope, T. M. H., Seghier, M. L., Parker Jones, O., Prejawa, S., Green, D. W., & Price, C. J. (2016). Four functionally distinct regions in the left supramarginal gyrus support word processing. *Cerebral Cortex, 26,* 4212–4226.

Pashek, G. V., & Holland, A. L. (1988). Evolution of aphasia in the first year post-onset. *Cortex, 24,* 411–423.

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

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10,* 437–442.

Pillai, J. J., & Zacà, D. (2011). Relative utility for hemispheric lateralization of different clinical fMRI activation tasks within a comprehensive language paradigm battery in brain tumor patients as assessed by both threshold-dependent and threshold-independent analysis methods. *NeuroImage, 54,* S136–S145.

Pillay, S. B., Stengel, B. C., Humphries, C., Book, D. S., & Binder, J. R. (2014). Cerebral localization of impaired phonological retrieval during rhyme judgment. *Annals of Neurology, 76,* 738–746.

Poldrack, R. A., Temple, E., Protopapas, A., Nagarajan, S., Tallal, P., Merzenich, M., & Gabrieli, J. D. E. (2001). Relations between the neural bases of dynamic auditory processing and phonological processing: Evidence from fMRI. *Journal of Cognitive Neuroscience, 13,* 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,* 15–35.

Price, C. J., Moore, C. J., Humphreys, G. W., & Wise, R. J. S. (1997). Segregating semantic from phonological processes during reading. *Journal of Cognitive Neuroscience, 9,* 727–733.

Price, C. J., Crinion, J., & Friston, K. J. (2006). Design and analysis of fMRI studies with neurologically impaired patients. *Journal of Magnetic Resonance Imaging*, 23, 816-826.

Pugh, K. R., Shaywitz, B. A., Shaywitz, S. E., Constable, R. T., Skudlarski, P., Fulbright, R. K., Bronen, R.A., Shankweiler, D.P., Katz, L., & Gore, J. C. (1996). Cerebral organization of component processes in reading. *Brain, 119,* 1221–1238.

Raboyeau, G., De Boissezon, X., Marie, N., Balduyck, S., Puel, M., Bézy, C., Démonet, J.F., & Cardebat, D. (2008). Right hemisphere activation in recovery from aphasia lesion effect or function recruitment? *Neurology*, *70*, 290-298.

Rau, S., Fesl, G., Bruhns, P., Havel, P., Braun, B., Tonn, J. C., & Ilmberger, J. (2007). Reproducibility of activations in Broca area with two language tasks: a functional MR imaging study. *American Journal of Neuroradiology*, *28*, 1346-1353.

Reppen, R., Ide, N., Suderman, K. (2005): American National Corpus (ANC) Second Release. Philadelphia: Linguistic Data Consortium.

Rombouts, S. A., Barkhof, F., Hoogenraad, F. G., Sprenger, M., Valk, J., & Scheltens, P. (1997). Test-retest analysis with functional MR of the activated area in the human visual cortex. *American Journal of Neuroradiology, 18,* 1317–1322.

Rosen, H. J., Petersen, S. E., Linenweber, M. R., Snyder, A. Z., White, D. A., Chapman, L., Dromerick, A.W., Fiez, J.A., & Corbetta, M. (2000). Neural correlates of recovery from aphasia after damage to left inferior frontal cortex. *Neurology*, *55*, 1883-1894.

Saur, D., Lange, R., Baumgaertner, A., Schraknepper, V., Willmes, K., Rijntjes, M., Weiller, C. (2006): Dynamics of language reorganization after stroke. *Brain*, *129,* 1371–1384.

Saur, D., Hartwigsen, G. (2012): Neurobiology of language recovery after stroke: lessons from neuroimaging studies. *Arch Phys Med Rehabil*, *93*, S15–S25.

Seghier, M. L., Lazeyras, F., Pegna, A. J., Annoni, J.-M., Zimine, I., Mayer, E., Michel, C.M., & Khateb, A. (2004). Variability of fMRI activation during a phonological and semantic language task in healthy subjects. *Human Brain Mapping, 23,* 140–155.

Seghier, M. L., Kherif, F., Josse, G., & Price, C. J. (2011). Regional and hemispheric determinants of language laterality: Implications for preoperative fMRI. *Human Brain Mapping, 32,* 1602–1614.

Seghier, M. L., Patel, E., Prejawa, S., Ramsden, S., Selmer, A., Lim, L., Browne, R., Rae, J., Haigh, Z., Ezekiel, D., Hope, T. M., Leff, A.P., & Price, C.J. (2016). The PLORAS database: a data repository for predicting language outcome and recovery after stroke. *NeuroImage*, *124,* 1208-1212.

Shaywitz, B. A., Shaywltz, S. E., Pugh, K. R., Constable, R. T., Skudlarski, P., Fulbright, R. K., Bronen, R.A., Fletcher, J.M., Shankweiler, D.P., Katz, L., & Gore, J. C. (1995). Sex differences in the functional organization of the brain for language. *Nature, 373,* 607–609.

Sliwinska, M. W., Khadilkar, M., Campbell-Ratcliffe, J., Quevenco, F., & Devlin, J. T. (2012). Early and sustained supramarginal gyrus contributions to phonological processing. *Frontiers in Psychology, 3,* 161.

Smith, E. E., Jonides, J., Marshuetz, C., & Koeppe, R. A. (1998). Components of verbal working memory: evidence from neuroimaging. *Proceedings of the National Academy of Sciences*, *95*, 876-882.

Spitsyna, G., Warren, J. E., Scott, S. K., Turkheimer, F. E., & Wise, R. J. (2006). Converging language streams in the human temporal lobe. *Journal of Neuroscience*, *26*, 7328-7336.

Springer, J. A., Binder, J. R., Hammeke, T. A., Swanson, S. J., Frost, J. A., Bellgowan, P. S. F., Brewer, C.C., Perry, H.M., Morris, G.L., & Mueller, W. M. (1999). Language dominance in neurologically normal and epilepsy subjects. *Brain, 122,* 2033–2046.

Stromswold, K., Caplan, D., Alpert, N., & Rauch, S. (1996). Localization of syntactic comprehension by positron emission tomography. *Brain and language*, *52*, 452-473.

Taylor, J. S. H., Rastle, K., & Davis, M. H. (2013). Can cognitive models explain brain activation during word and pseudoword reading? A meta-analysis of 36 neuroimaging studies. *Psychological Bulletin*, 139, 766–791.

Taylor, J. S. H., Rastle, K., & Davis, M. H. (2014). Interpreting response time effects in functional imaging studies. *NeuroImage*, *99*, 419–433.

Thompson, C.K., den Ouden, D.-B. (2008): Neuroimaging and recovery of language in aphasia. *Curr Neurol Neurosci Rep*, *8*, 475–483.

Trojano, L., Grossi, D., Linden, D. E. J., Formisano, E., Hacker, H., Zanella, F. E., Goebel, R., & Di Salle, F. (2000). Matching two imagined clocks: the functional anatomy of spatial analysis in the absence of visual stimulation. *Cerebral Cortex, 10,* 473–481.

Turkeltaub, P.E., Messing, S., Norise, C., Hamilton, R.H. (2011): Are networks for residual language function and recovery consistent across aphasic patients? *Neurology*, *76*, 1726–1734.

Twomey, T., Waters, D., Price, C. J., Kherif, F., Woll, B., & MacSweeney, M. (2015). Identification of the regions involved in phonological assembly using a novel paradigm. *Brain and Language, 150,* 45–53.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., & Joliot, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage, 15,* 273–289.

Tzourio-Mazoyer, N., Petit, L., Razafimandimby, A., Crivello, F., Zago, L., Jobard, G., Joliot, M., Mellet, E., & Mazoyer, B. (2010). Left hemisphere lateralization for language in right-handers is controlled in part by familial sinistrality, manual preference strength, and head size. *Journal of Neuroscience, 30,* 13314–13318.

Vallar, G., Bisiach, E., Cerizza, M., & Rusconi, M. L. (1988). The role of the left hemisphere in decision-making. *Cortex, 24,* 399–410.

Warburton, E., Price, C. J., Swinburn, K., & Wise, R. J. (1999). Mechanisms of recovery from aphasia: evidence from positron emission tomography studies. *Journal of Neurology, Neurosurgery & Psychiatry*, *66*, 155-161.

Weiller, C., Isensee, C., Rijntjes, M., Huber, W., Müller, S., Bier, D., Dutschka, K., Woods, R.P., Noth, J., & Diener, H. C. (1995). Recovery from Wernicke's aphasia: a positron emission tomographic study. *Annals of neurology*, *37*, 723 732.

Wilke, M., & Lidzba, K. (2007). LI-tool: A new toolbox to assess lateralization in functional MR-data. *Journal of Neuroscience Methods, 163,* 128–136.

Wilson, S. M., Brambati, S. M., Henry, R. G., Handwerker, D. A., Agosta, F., Miller, B. L., Wilkins, D.P., Ogar, J.M., & Gorno-Tempini, M. L. (2009). The neural basis of surface dyslexia in semantic dementia. *Brain, 132,* 71–86.

Wilson, S. M., DeMarco, A. T., Henry, M. L., Gesierich, B., Babiak, M., Miller, B. L., & Gorno-Tempini, M. L. (2016). Variable disruption of a syntactic processing network in primary progressive aphasia. *Brain*, 139, 2994-3006.

Wilson, S. M., Bautista, A., Yen, M., Lauderdale, S., & Eriksson, D. K. (2017). Validity and reliability of four language mapping paradigms. *NeuroImage: Clinical, 16,* 399–408.

Wilson, S. M., Yen, M., & Eriksson, D. K. (2018). An adaptive semantic matching paradigm for reliable and valid language mapping in individuals with aphasia. *Human Brain Mapping, 39*, 3285-32307.

Wilson, S. M., Eriksson, D. K., Schneck, S. M., & Lucanie, J. M. (2018). A quick aphasia battery for efficient, reliable, and multidimensional assessment of language function. *PLoS One, 13,* e0192773.

Winkler, A. M., Ridgway, G. R., Webster, M. A., Smith, S. M., & Nichols, T. E. (2014). Permutation inference for the general linear model. *NeuroImage, 92,* 381–397.

Winhuisen, L., Thiel, A., Schumacher, B., Kessler, J., Rudolf, J., Haupt, W. F., & Heiss, W. D. (2005). Role of the contralateral inferior frontal gyrus in recovery of language function in poststroke aphasia: a combined repetitive transcranial magnetic stimulation and positron emission tomography study. *Stroke*, *36*, 1759-1763.

Winhuisen, L., Thiel, A., Schumacher, B., Kessler, J., Rudolf, J., Haupt, W. F., & Heiss, W. D. (2007). The right inferior frontal gyrus and poststroke aphasia: a follow-up investigation. *Stroke*, *38*, 1286-1292.

Worsley, K. J., Liao, C. H., Aston, J., Petre, V., Duncan, G. H., Morales, F., & Evans, A. C. (2002). A general statistical analysis for fMRI data. *NeuroImage, 15,* 1–15.

Xing, S., Lacey, E. H., Skipper-Kallal, L. M., Jiang, X., Harris-Love, M. L., Zeng, J., & Turkeltaub, P. E. (2015). Right hemisphere grey matter structure and language outcomes in chronic left hemisphere stroke. *Brain*, *139*, 227-241.

Yagata, S. A., Yen, M., McCarron, A., Bautista, A., Lamair-Orosco, G., & Wilson, S. M. (2016). Rapid recovery from aphasia after infarction of Wernicke's area. *Aphasiology, 31,* 951–980.

Yeo, B. T. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., Roffman, J. L., Smoller, J. W., Zöllei, L., Polimeni, J. R., Fischl, B., Liu, H., & Buckner, R. L. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology*, *106*, 1125–1165.

Zacà, D., Nickerson, J. P., Deib, G., & Pillai, J. J. (2012). Effectiveness of four different clinical fMRI paradigms for preoperative regional determination of language lateralization in patients with brain tumors. *Neuroradiology, 54,* 1015–1025.