Adaptive paradigms for mapping phonological regions in individual participants

Melodie Yen[1,*], **Andrew T. DeMarco**[2], and **Stephen M. Wilson**[1]

[1]Department of Hearing and Speech Sciences, Vanderbilt University Medical Center, Nashville, TN, USA

[2]Department of Neurology, Georgetown University Medical Center, Washington, DC, USA

Abstract

Phonological encoding depends on left-lateralized regions in the supramarginal gyrus and the ventral precentral gyrus. Localization of these phonological regions in individual participants—including individuals with language impairments—is important in several research and clinical contexts. To localize these regions, we developed two paradigms that load on phonological encoding: a rhyme judgment task and a syllable counting task. Both paradigms relied on an adaptive staircase design to ensure that each individual performed each task at a similarly challenging level. The goal of this study was to assess the validity and reliability of the two paradigms, in terms of their ability to consistently produce left-lateralized activations of the supramarginal gyrus and ventral precentral gyrus in neurologically normal individuals with presumptively normal language localization. Sixteen participants were scanned with fMRI as they performed the rhyme judgment paradigm, the syllable counting paradigm, and an adaptive semantic paradigm that we have described previously. We found that the rhyme and syllable paradigms both yielded left-lateralized supramarginal and ventral precentral activations in the majority of participants. The rhyme paradigm produced more lateralized and more reliable activations, and so should be favored in future applications. In contrast, the semantic paradigm did not reveal supramarginal or precentral activations in most participants, suggesting that the recruitment of these regions is indeed driven by phonological encoding, not language processing in general. In sum, the adaptive rhyme judgment paradigm was effective in localizing left-lateralized phonological encoding regions in individual participants, and, in conjunction with the adaptive semantic paradigm, can be used to map individual language networks.

Introduction

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

*Corresponding author. Department of Hearing and Speech Sciences, Vanderbilt University Medical Center, 1215 21st Ave S, MCE 8310, Nashville, TN 37232. melodie.yen@vanderbilt.edu.

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, in the context of investigating functional reorganization of language processing in recovery from aphasia, it would be beneficial to be able to identify phonological regions. 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. The semantic 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 depend on 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 depended on 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 imaging 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) native 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 (MMSE; Folstein et al., 1975) and the Quick Aphasia Battery (QAB; 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 perfomed 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. ΘδႪႥ- ΘδႪႥ), while the other half were not identical (e.g. ΘδႪႥ-Ⴥ ႱႱ). 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 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.

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). 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.

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 = $2.3 \times 2.3 \times 3.5$ 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 $LI = (V_{Left} - V_{Right}) / (V_{Left} + V_{Right})$, 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; a repeated measures ANOVA showed a significant interaction of paradigm by condition for accuracy ($F(2, 30) = 19.18$, $p < .0001$) (Fig. 1A). Follow-up 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 differed across paradigms and conditions (Fig. 1B). Although it is informative to see the range of mean difficulty levels at which items were presented for each participant, we did not examine this apparent interaction statistically, because the seven difficulty levels were not inherently equivalent across conditions or across paradigms.

Finally, reaction times too showed a significant interaction of paradigm by condition ($F(2, 30) = 32.14$, $p < .0001$) (Fig. 1C). 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.

Neuroimaging results

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

differences were statistically significant. Coordinates and statistical details for all activated regions are provided in Table 3.

To determine whether differences between paradigms might reflect confounds of accuracy or reaction time, additional analyses were carried out in which differences between paradigms in accuracy or reaction time relative to the perceptual control condition were included as covariates. No regions were significantly correlated with any of these covariates, reducing the likelihood that the differences observed between paradigms were secondary to confounds of accuracy or reaction time.

Sensitivity—Sensitivity was compared across paradigms in each ROI (Fig. 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 2,000 mm^3 extent: 13 out of 16 on the rhyme paradigm and 13 out of 16 on the syllable counting paradigm. 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 2,000 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 ($\chi^2(2) = 24.94$, $p < .0001$).

The possibility was considered that supramarginal and/or ventral precentral activations for the rhyme or syllables paradigms might reflect the lower accuracy of participants on those language tasks compared to their perceptual control conditions. If activations were related to the greater difficulty of the language tasks, or to the commission or monitoring of errors, then positive correlations would be predicted across participants between error rate differences and activation extent. No such positive correlations were observed (supramarginal rhyme: $r = -0.13$, $p = .62$; supramarginal syllables: $r = -0.18$, $p = .50$; precentral rhyme: $r = -0.14$, $p = .61$; precentral syllables: $r = -0.05$, $p = .87$).

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 extensive 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 ($|t(15)| = 2.13$, $p = .051$).

Lateralization—Lateralization was compared across paradigms in each ROI (Fig. 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, $p = .018$). 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; $p = .18$). Note that lateralization was not assessed for the semantic paradigm in these regions becasue 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 $p = .0646$).

Reliability—In the split-half analyses of test-retest reproducibility, the Dice coefficient of similarity differed across the three paradigms ($F(1, 30) = 6.57$, $p = .0057$) (Fig. 5). 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 ($|t(15)| = 1.29$, $p = .22$).

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. 6. 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.

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. 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 particpant 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.

Conclusion

Despite these limitations, the adaptive rhyme judgment task in particular shows considerable promise as a paradigm capable of mapping known phonological encoding regions in the left supramarginal gyrus and left ventral precentral gyrus in individual participants. The present findings establish the validity and reliability of the paradigm in neurologically normal controls, and suggest that future studies are warranted to address the utility of the paradigm in presurgical patients and individuals with aphasia. By using the adaptive rhyme judgment and adaptive semantic paradigms in conjunction, it should be possible to map domain-specific language regions in individual participants, allowing us to go beyond the concept of simply mapping undifferentiated "eloquent cortex". This offers clear advantages for

presurgical language mapping, studies of neuroplasticity in recovery from aphasia, and studies that use functional localizers to probe the functionality of different nodes of the language network.

Acknowledgments

References

Ashburner J, & Friston KJ (2005). Unified segmentation. NeuroImage, 26, 839–851. [PubMed: 15955494]

Baciu M, Kahane P, Minotti L, Charnallet A, David D, Le Bas JF, & 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. [PubMed: 11679302]

Baciu MV, Watson JM, Maccotta L, McDermott KB, Buckner RL, Gilliam FG, & Ojemann JG (2005). Evaluating functional MRI procedures for assessing hemispheric language dominance in neurosurgical patients. Neuroradiology, 47, 835–844. [PubMed: 16142480]

Baldo JV, & Dronkers NF (2006). The role of inferior parietal and inferior frontal cortex in working memory. Neuropsychology, 20, 529–538. [PubMed: 16938015]

Beckmann CF, & Smith SM (2004). Probabilistic independent component analysis for functional magnetic resonance imaging. IEEE Transactions on Medical Imaging, 23, 137–152. [PubMed: 14964560]

Billingsley RL, McAndrews MP, Crawley AP, & Mikulis DJ (2001). Functional MRI of phonological and semantic processing in temporal lobe epilepsy. Brain, 124, 1218–1227. [PubMed: 11353737]

Binder JR, Desai RH, Graves WW, & Conant LL (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. Cerebral Cortex, 19, 2767–2796. [PubMed: 19329570]

Binder JR, Medler DA, Desai R, Conant LL, & Liebenthal E (2005). Some neurophysiological constraints on models of word naming. NeuroImage, 27, 677–693. [PubMed: 15921937]

Binder JR, Swanson SJ, Hammeke TA, & Sabsevitz DS (2008). A comparison of five fMRI protocols for mapping speech comprehension systems. Epilepsia, 49, 1980–1997. [PubMed: 18513352]

Booth JR, Burman DD, Meyer JR, Gitelman DR, Parrish TB, & Mesulam MM (2002). Functional anatomy of intra- and cross-modal lexical tasks. NeuroImage, 16, 7–22. [PubMed: 11969313]

Booth JR, Lu D, Burman DD, 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. [PubMed: 16427033]

Bradshaw AR, Thompson PA, Wilson AC, Bishop DVM, & Woodhead ZVJ (2017). Measuring language lateralisation with different language tasks: a systematic review. PeerJ, 5, e3929. [PubMed: 29085748]

Brainard DH (1997). The psychophysics toolbox. Spatial Vision, 10, 433–436. [PubMed: 9176952]

Braun AR, Varga M, Stager S, Schulz G, Selbie S, Maisog JM, … Ludlow CL (1997). Altered patterns of cerebral activity during speech and language production in developmental stuttering: an H215O positron emission tomography study. Brain, 120, 761–784. [PubMed: 9183248]

Burton MW, LoCasto PC, Krebs-Noble D, & Gullapalli RP (2005). A systematic investigation of the functional neuroanatomy of auditory and visual phonological processing. NeuroImage, 26, 647–661. [PubMed: 15955475]

Clements AM, Rimrodt SL, Abel JR, Blankner JG, Mostofsky SH, Pekar JJ, Denckla MB, & Cutting LE (2006). Sex differences in cerebral laterality of language and visuospatial processing. Brain and Language, 98, 150–158. [PubMed: 16716389]

Coltheart M (1981). The MRC psycholinguistic database. Quarterly Journal of Experimental Psychology Section A, 33, 497–505.

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. [PubMed: 17339089]

Cox RW (1996). AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Computers and Biomedical Research, 29, 162–173. [PubMed: 8812068]

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

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. [PubMed: 24062451]

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. [PubMed: 20410363]

Folstein MF, Folstein SE, & McHugh PR (1975). "Mini-mental state." Journal of Psychiatric Research, 12, 189–198. [PubMed: 1202204]

Fridriksson J, Kjartansson O, Morgan PS, Hjaltason H, Magnusdottir S, Bonilha L, & Rorden C (2010). Impaired speech repetition and left parietal lobe damage. Journal of Neuroscience, 30, 11057–11061. [PubMed: 20720112]

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. [PubMed: 27956600]

Gaillard WD, Berl MM, Moore EN, Ritzl EK, Rosenberger LR, Weinstein SL, Conry JA, Pearl PL, Ritter FF, Sato S, Vezina LG, Vaidya CJ, Wiggs E, Fratalli C, Risse G, Ratner NB, Gioia G, & Theodore WH (2007). Atypical language in lesional and nonlesional complex partial epilepsy. Neurology, 69, 1761–1771. [PubMed: 17967992]

García-Pérez MA (1998). Forced-choice staircases with fixed step sizes: Asymptotic and small-sample properties. Vision Research, 38, 1861–1881. [PubMed: 9797963]

Geranmayeh F, Brownsett SLE, & Wise RJS (2014). Task-induced brain activity in aphasic stroke patients: What is driving recovery? Brain, 137, 2632–2648. [PubMed: 24974382]

Geva S, Jones PS, Crinion JT, Price CJ, Baron J-C, & Warburton EA (2011). The neural correlates of inner speech defined by voxel-based lesion-symptom mapping. Brain, 134, 3071–3082. [PubMed: 21975590]

Gitelman DR, Nobre AC, Sonty S, Parrish TB, & Mesulam M-M (2005). Language network specializations: An analysis with parallel task designs and functional magnetic resonance imaging. NeuroImage, 26, 975–985. [PubMed: 15893473]

Gough PM, Nobre AC, & Devlin JT (2005). Dissociating linguistic processes in the left inferior frontal cortex with transcranial magnetic stimulation. Journal of Neuroscience, 25, 8010–8016. [PubMed: 16135758]

Grabski K, Tremblay P, Gracco VL, 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. [PubMed: 23542585]

Gross WL, & Binder JR (2014). Alternative thresholding methods for fMRI data optimized for surgical planning. NeuroImage, 84, 554–561. [PubMed: 24021837]

Hartwigsen G, Price CJ, Baumgaertner A, Geiss G, Koehnke M, Ulmer S, & Siebner HR (2010). The right posterior inferior frontal gyrus contributes to phonological word decisions in the healthy brain: Evidence from dual-site TMS. Neuropsychologia, 48, 3155–3163. [PubMed: 20600177]

Hartwigsen G, Weigel A, Schuschan P, Siebner HR, Weise D, Classen J, & Saur D (2016). Dissociating parieto-frontal networks for phonological and semantic word decisions: A condition-and-perturb TMS study. Cerebral Cortex, 26, 2590–2601. [PubMed: 25953770]

Ito S, Stuphorn V, Brown JW, & Schall JD (2003). Performance monitoring by the anterior cingulate cortex during saccade countermanding. Science, 302, 120–122. [PubMed: 14526085]

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. [PubMed: 14568445]

Kelly RE, Alexopoulos GS, Wang Z, Gunning FM, Murphy CF, Morimoto SS, Kanellopoulos D, Jia Z, Lim KO, & Hoptman MJ (2010). Visual inspection of independent components: Defining a procedure for artifact removal from fMRI data. Journal of Neuroscience Methods, 189, 233–245. [PubMed: 20381530]

Keuleers E, & Brysbaert M (2010). Wuggy: A multilingual pseudoword generator. Behavior Research Methods, 42, 627–633. [PubMed: 20805584]

Knecht S, Jansen A, Frank A, van Randenborgh J, Sommer J, Kanowski M, & Heinze HJ (2003). How atypical is atypical language dominance? NeuroImage, 18, 917–927. [PubMed: 12725767]

Laska AC, 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. [PubMed: 11350565]

Levelt WJM (1989). Speaking: from intention to articulation. Cambridge, MA: MIT Press.

Levelt WJM (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. [PubMed: 11698690]

Lurito JT, Kareken DA, Lowe MJ, Chen SHA, & Mathews VP (2000). Comparison of rhyming and word generation with FMRI. Human Brain Mapping, 10, 99–106. [PubMed: 10912589]

McDermott KB, Petersen SE, Watson JM, & Ojemann JG (2003). A procedure for identifying regions preferentially activated by attention to semantic and phonological relations using functional magnetic resonance imaging. Neuropsychologia, 41, 293–303. [PubMed: 12457755]

Mechelli A, Gorno-Tempini ML, & Price CJ (2003). Neuroimaging studies of word and pseudoword reading: consistencies, inconsistencies, and limitations. Journal of Cognitive Neuroscience, 15, 260–271. [PubMed: 12676063]

Mirman D, Chen Q, Zhang Y, Wang Z, Faseyitan OK, Coslett HB, & Schwartz MF (2015). Neural organization of spoken language revealed by lesion–symptom mapping. Nature Communications, 6, 6762.

Morrison MA, Churchill NW, Cusimano MD, Schweizer TA, Das S, & Graham SJ (2016). Reliability of task-based fMRI for preoperative planning: A test-retest study in brain tumor patients and healthy controls. PLoS One, 11, e0149547. [PubMed: 26894279]

Nichols TE, & Holmes AP (2002). Nonparametric permutation tests for functional neuroimaging: A primer with examples. Human Brain Mapping, 15, 1–25. [PubMed: 11747097]

Oberhuber M, Hope TMH, Seghier ML, Parker Jones O, Prejawa S, Green DW, & Price CJ (2016). Four functionally distinct regions in the left supramarginal gyrus support word processing. Cerebral Cortex, 26, 4212–4226.

Pashek GV, & Holland AL (1988). Evolution of aphasia in the first year post-onset. Cortex, 24, 411–423. [PubMed: 3191724]

Paulesu E, Frith CD, & Frackowiak RSJ (1993). The neural correlates of the verbal component of working memory. Nature, 362, 342–345. [PubMed: 8455719]

Pelli DG (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. Spatial Vision, 10, 437–442. [PubMed: 9176953]

Pillai JJ, & 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. [PubMed: 20380883]

Pillay SB, Stengel BC, Humphries C, Book DS, & Binder JR (2014). Cerebral localization of impaired phonological retrieval during rhyme judgment. Annals of Neurology, 76, 738–746. [PubMed: 25164766]

Poldrack RA, Temple E, Protopapas A, Nagarajan S, Tallal P, Merzenich M, & Gabrieli JDE (2001). Relations between the neural bases of dynamic auditory processing and phonological processing: Evidence from fMRI. Journal of Cognitive Neuroscience, 13, 687–697. [PubMed: 11506664]

Poldrack RA, Wagner AD, Prull MW, Desmond JE, Glover GH, & Gabrieli JDE (1999). Functional specialization for semantic and phonological processing in the left inferior prefrontal cortex. NeuroImage, 10, 15–35. [PubMed: 10385578]

Price CJ, Crinion J, & Friston KJ (2006). Design and analysis of fMRI studies with neurologically impaired patients. Journal of Magnetic Resonance Imaging, 23, 816–826. [PubMed: 16649208]

Price CJ, Moore CJ, Humphreys GW, & Wise RJS (1997). Segregating semantic from phonological processes during reading. Journal of Cognitive Neuroscience, 9, 727–733. [PubMed: 23964595]

Pugh KR, Shaywitz BA, Shaywitz SE, Constable RT, Skudlarski P, Fulbright RK, Bronen RA, Shankweiler DP, Katz L, & Gore JC (1996). Cerebral organization of component processes in reading. Brain, 119, 1221–1238. [PubMed: 8813285]

Rombouts SA, Barkhof F, Hoogenraad FG, 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. [PubMed: 9282862]

Seghier ML, Kherif F, Josse G, & Price CJ (2011). Regional and hemispheric determinants of language laterality: Implications for preoperative fMRI. Human Brain Mapping, 32, 1602–1614. [PubMed: 20814960]

Seghier ML, Lazeyras F, Pegna AJ, Annoni J-M, Zimine I, Mayer E, Michel CM, & Khateb A (2004). Variability of fMRI activation during a phonological and semantic language task in healthy subjects. Human Brain Mapping, 23, 140–155. [PubMed: 15449358]

Shaywitz BA, Shaywltz SE, Pugh KR, Constable RT, Skudlarski P, Fulbright RK, Bronen RA, Fletcher JM, Shankweiler DP, Katz L, & Gore JC (1995). Sex differences in the functional organization of the brain for language. Nature, 373, 607–609. [PubMed: 7854416]

Sliwinska MW, Khadilkar M, Campbell-Ratcliffe J, Quevenco F, & Devlin JT (2012). Early and sustained supramarginal gyrus contributions to phonological processing. Frontiers in Psychology, 3, 161. [PubMed: 22654779]

Springer JA, Binder JR, Hammeke TA, Swanson SJ, Frost JA, Bellgowan PSF, Brewer CC, Perry HM, Morris GL, & Mueller WM (1999). Language dominance in neurologically normal and epilepsy subjects. Brain, 122, 2033–2046. [PubMed: 10545389]

Taylor JSH, Rastle K, & Davis MH (2013). Can cognitive models explain brain activation during word and pseudoword reading? A meta-analysis of 36 neuroimaging studies. Psychological Bulletin, 139, 766–791. [PubMed: 23046391]

Trojano L, Grossi D, Linden DEJ, Formisano E, Hacker H, Zanella FE, 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. [PubMed: 10847597]

Twomey T, Waters D, Price CJ, 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. [PubMed: 26335996]

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. [PubMed: 11771995]

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. [PubMed: 20926657]

Wilke M, & Lidzba K (2007). LI-tool: A new toolbox to assess lateralization in functional MR-data. Journal of Neuroscience Methods, 163, 128–136. [PubMed: 17386945]

Wilson M (1988). MRC psycholinguistic database: Machine-usable dictionary, version 2.00. Behavior Research Methods, Instruments, & Computers, 20, 6–10.

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

Wilson SM, Brambati SM, Henry RG, Handwerker DA, Agosta F, Miller BL, Wilkins DP, Ogar JM, & Gorno-Tempini ML (2009). The neural basis of surface dyslexia in semantic dementia. Brain, 132, 71–86. [PubMed: 19022856]

Wilson SM, Eriksson DK, Schneck SM, & Lucanie JM (2018). A quick aphasia battery for efficient, reliable, and multidimensional assessment of language function. PLoS One, 13, e0192773. [PubMed: 29425241]

Wilson SM, Yen M, & Eriksson DK (2018). An adaptive semantic matching paradigm for reliable and valid language mapping in individuals with aphasia. Human Brain Mapping, in press.

Winkler AM, Ridgway GR, Webster MA, Smith SM, & Nichols TE (2014). Permutation inference for the general linear model. NeuroImage, 92, 381–397. [PubMed: 24530839]

Worsley KJ, Liao CH, Aston J, Petre V, Duncan GH, Morales F, & Evans AC (2002). A general statistical analysis for fMRI data. NeuroImage, 15, 1–15. [PubMed: 11771969]

Yagata SA, Yen M, McCarron A, Bautista A, Lamair-Orosco G, & Wilson SM (2016). Rapid recovery from aphasia after infarction of Wernicke's area. Aphasiology, 31, 951–980. [PubMed: 29051682]

Zacà D, Nickerson JP, Deib G, & Pillai JJ (2012). Effectiveness of four different clinical fMRI paradigms for preoperative regional determination of language lateralization in patients with brain tumors. Neuroradiology, 54, 1015–1025. [PubMed: 22744798]
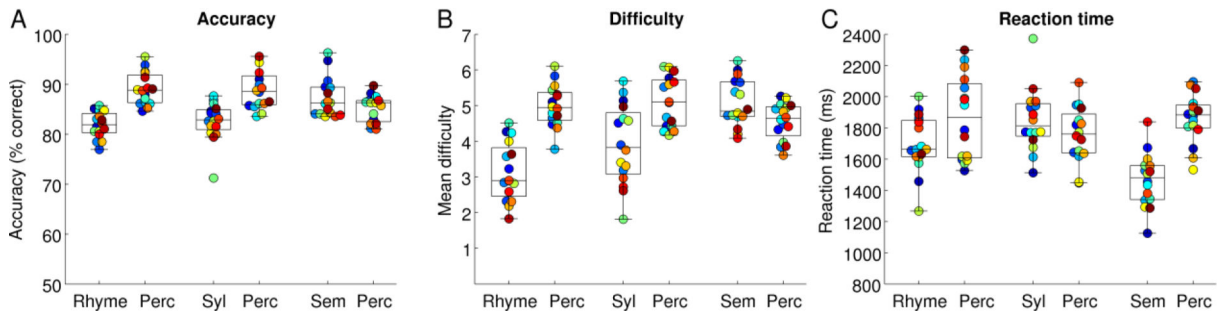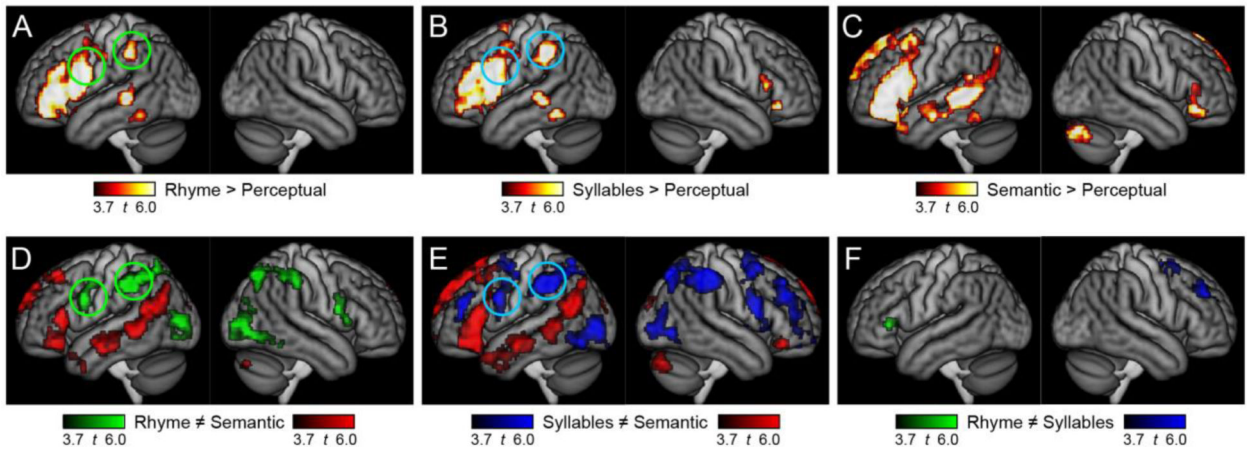
Figure 1.
Behavioral results. Each participant is denoted with a unique color. (A) Accuracy. (B) Difficulty level of items presented. (C) Reaction time. Rhyme = Rhyme judgment; Syl = Syllable counting; Sem = Semantic matching; Perc = Perceptual control task.
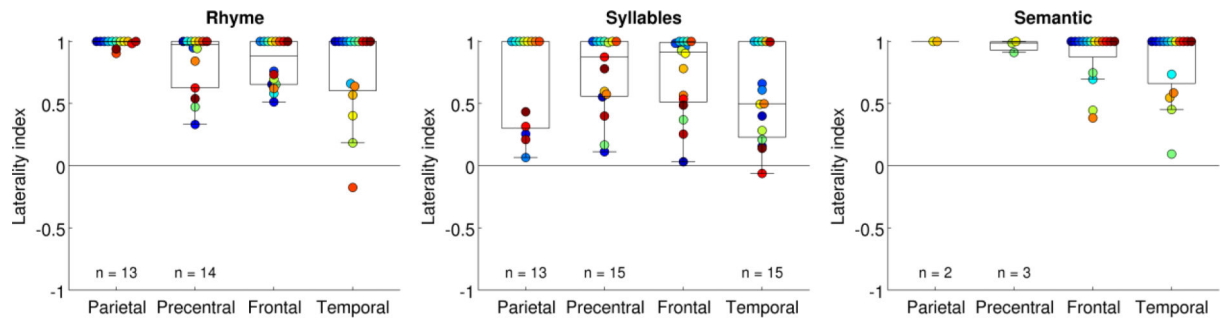
Figure 2.
Activation maps for group contrasts. The top row shows each language task compared to its perceptual control task: (A) Rhyme > Perceptual; (B) Syllables > Perceptual; (C) Semantic > Perceptual. The bottom row shows pairwise contrasts between the language paradigms: (D) Rhyme   Semantic; (E) Syllables   Semantic; (F) Rhyme   Syllables.
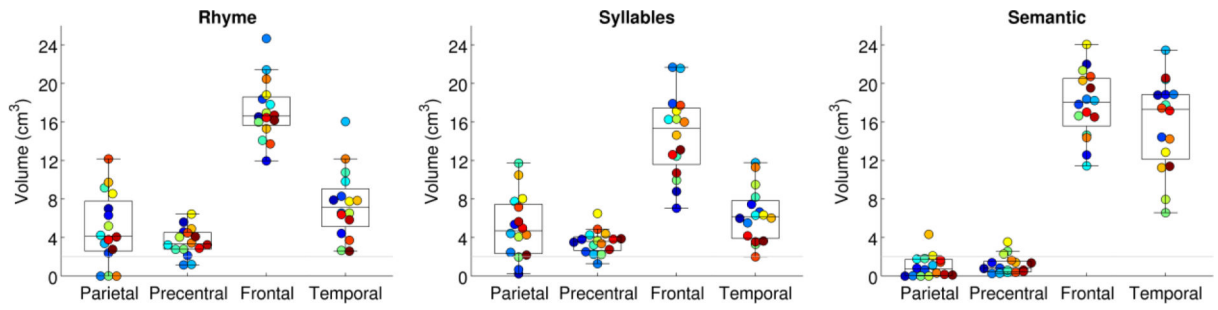
Figure 3.
Extent of activation for each paradigm in each region of interest. The gray horizontal line indicates the 2,000 mm^3 extent of activation which was our threshold for considering a region activated. Each participant is denoted with a unique color, matching Fig. 1.
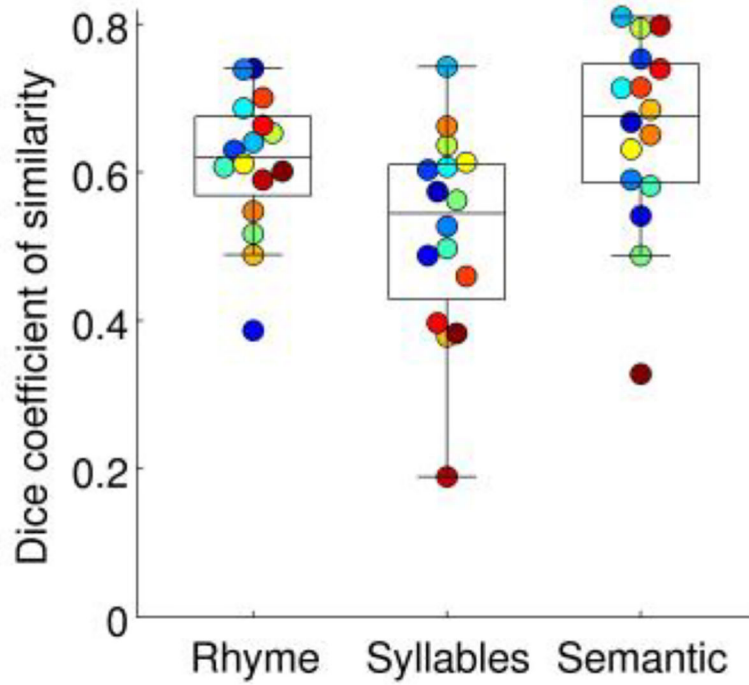
Figure 4.
Lateralization of activation for each paradigm in each region of interest. Each participant is denoted with a unique color, matching previous figures.

Figure 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.
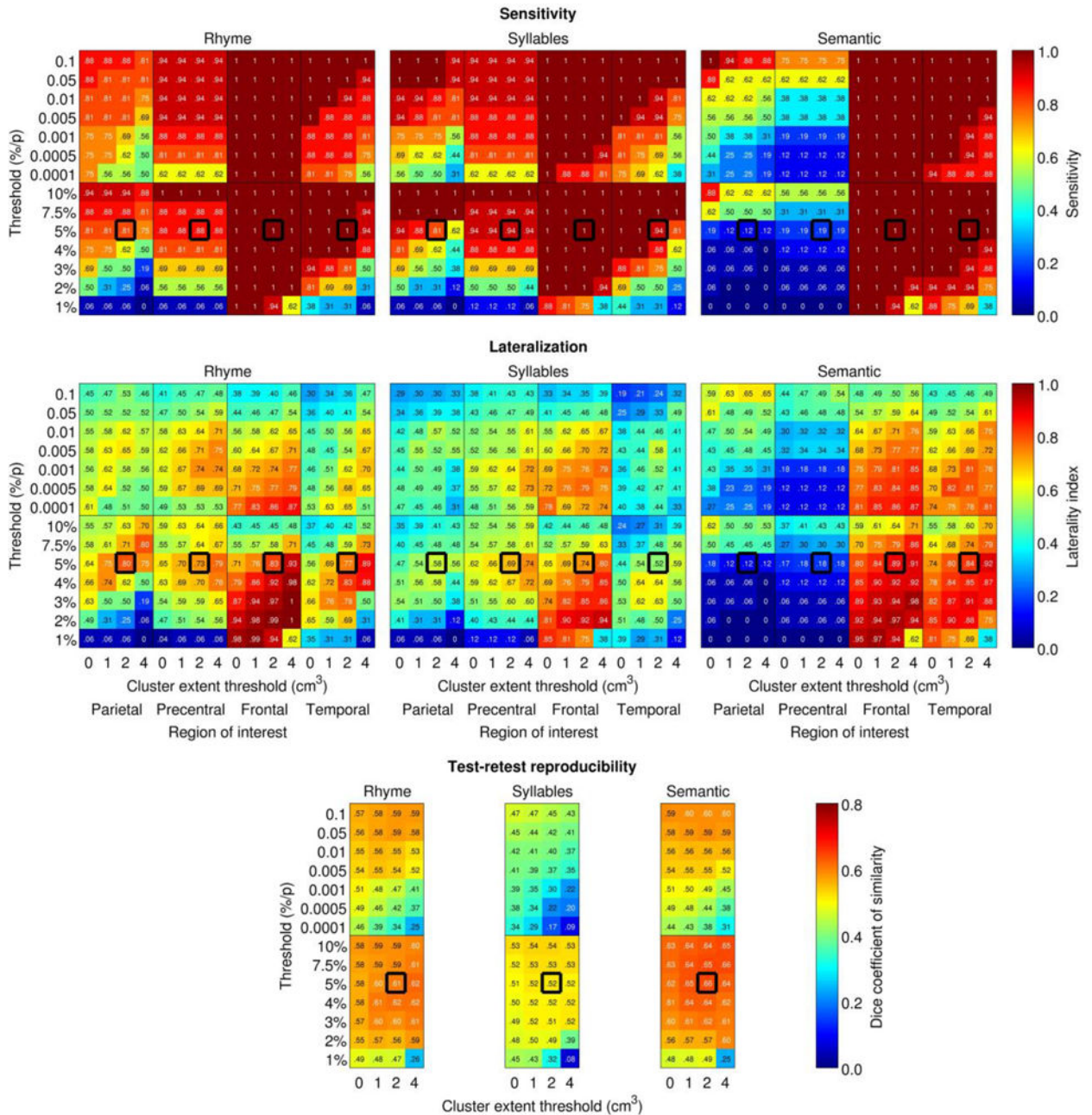
Figure 6.
Impact of analysis parameters on sensitivity, laterality, and reliability. Thick black outlines denote the *a priori* analysis parameters. Sensitivity for detection, lateralization indices, and Dice coefficients are plotted as a function of absolute and relative voxelwise thresholds (*y* axes), 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.

Table 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*	ꝢჂჃ ꝢჂჃ	ᲂᲧᲣᲃᲣᲣᲤ ᲃᲔᲤᲔᲥ2
2	3.77 ± 0.43	2.00 ± 0.95	1.63 ± 0.81	*scaw* *praw*	*pake* *pape*	ꙆꙆჂᲥᲥ ꙆꙆჂᲥᲥ	ᲇᲣᲃᲣᲣ ᲗᲇᲣᲣᲣ
3	5.40 ± 0.97	3.00 ± 1.00	2.16 ± 0.75	*veery* *feary*	*shofy* *sheffy*	ᲇ ᲃᲤᲣ ᲇ ᲃᲤᲣ	²ᲥᲣ²ᲣᲧᲒᲔ ²ᲥᲣ²ᲣᲒᲔᲃ
4	5.50 ± 0.78	3.25 ± 1.29	2.75 ± 0.97	*lameed* *abede*	*renerve* *setarve*	ᲒᲘᲃᲃᲘ ᲒᲘᲃᲃᲘ	ᲃᲘ²ᲣᲤᲧ 2ᲃ²ᲣᲤᲧ
5	7.03 ± 1.04	3.40 ± 1.14	3.05 ± 1.03	*loschito* *fiuspeeto*	*lodacco* *topucko*	ᲃᲃᲣᲣᲥᲤ ᲃᲃᲣᲣᲥᲤ	ᲃ2²Ξ²ᒼ ²2²Ξ²ᒼ
6	9.45 ± 1.07	4.60 ± 1.28	4.27 ± 1.08	*cupadimote* *recadimoat*	*vermissible* *vanisibe*	ᲣᲣᲘᲘᲘᲬᲔᲤᲥ ᲣᲣ²ᲔᲔᲤᲥ	ᲧᲣᲘᲘᲘᲬᲔᲤᲤ ᲧᲥᲔᲘᲔᲔᲤ
7	11.58 ± 1.58	5.28 ± 1.31	4.92 ± 1.22	*cypermollicle* *waiperbollical*	*devapitulate* *rumipostulate*	ᲬᲥ²ᲃᲃᲘᲘ ᲬᲥ²ᲃᲃᲘᲘ	ᲘᲘᲘ Ꝋ

Table 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*	ꝷꙨꙅꝸ ꝷꙨꙅꝸ	ꙨΨℲꝸꙡⱮΦ ꝸŒΦKꙅ
2	3 or 4 letters 1 or 2 syllables	3 or 4 letters 1 or 2 syllables	*gock* *tiss*	*sork* *boma*	ΘΘꝸᒣᒣ ΘΘꝸᒣᒣ	ꟻΣꝸΘ ΘꙅΘΘꟻ
3	4 or 5 letters 1 or 2 syllables	4 or 5 letters 1 or 2 syllables	*lapo* *pany*	*namp* *ambus*	Ⴈ ꝸΦꙊ Ⴈ ꝸΦꙊ	ꙆᒣꙆΨŒ ꙅᒣꙆꙆŒꙆ
4	5 or 6 letters 1 or 2 syllables	5 or 6 letters 1 or 2 syllables	*tarmo* *vegar*	*blash* *ragle*	ŒΘꙆꙆꞪ ŒΘꙆꙆꞪ	ꝸΘꙆΦΨ ꙅꙆꙆΦΨ
5	6 or 7 letters 2 or 3 syllables	6 or 7 letters 2 or 3 syllables	*banlem* *terrow*	*esbort* *kisonic*	ꝸꝸꙊꙊꓘΦ ꝸꝸꙊꙊꓘΦ	ꝸꙅꙆΞꙆ ꙆꙅꙆΞꙆ
6	7 or 8 letters 2 or 3 syllables	7 or 8 letters 2 or 3 syllables	*opunate* *fundacy*	*displert* *pomable*	ꙡꙆΞΞΦꙆ ꙡꙆΞΞΦꙆ	ΨꙡΘℲꓘΞ ΨꓘΘŒꓘΞ
7	8 or 9 letters 3 or 4 syllables	8 or 9 letters 3 or 4 syllables	*uncament* *nanculous*	*bepidity* *delibance*	ℲꙆꙆΞꙆꞪ ℲꙆꙆΞꙆꞪ	ΘℲꓠŒ Θ ꓠŒ

Table 3

Functional activations for group contrasts

Brain region	MNI coordinates			Extent (mm³)	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

Brain region	MNI coordinates			Extent (mm^3)	Max t	p
	x	y	z			
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