Optimally Efficient Neural Systems for Processing Spoken Language

Jie Zhuang[1], Lorraine K. Tyler[1], Billi Randall[1], Emmanuel A. Stamatakis[1,2] and William D. Marslen-Wilson[1,3]

[1]Department of Psychology, University of Cambridge, Cambridge CB2 3EB, UK, [2]Division of Anaesthesia, School of Clinical Medicine, University of Cambridge, Cambridge CB2 2QQ, UK and [3]MRC Cognition and Brain Sciences Unit, Cambridge CB2 7EF, UK

Address correspondence to William D. Marslen-Wilson, Department of Psychology, University of Cambridge, Cambridge CB2 3EB, UK. Email: wdm10@cam.ac.uk

Cognitive models claim that spoken words are recognized by an optimally efficient sequential analysis process. Evidence for this is the finding that nonwords are recognized as soon as they deviate from all real words (Marslen-Wilson 1984), reflecting continuous evaluation of speech inputs against lexical representations. Here, we investigate the brain mechanisms supporting this core aspect of word recognition and examine the processes of competition and selection among multiple word candidates. Based on new behavioral support for optimal efficiency in lexical access from speech, a functional magnetic resonance imaging study showed that words with later nonword points generated increased activation in the left superior and middle temporal gyrus (Brodmann area [BA] 21/22), implicating these regions in dynamic sound-meaning mapping. We investigated competition and selection by manipulating the number of initially activated word candidates (competition) and their later drop-out rate (selection). Increased lexical competition enhanced activity in bilateral ventral inferior frontal gyrus (BA 47/45), while increased lexical selection demands activated bilateral dorsal inferior frontal gyrus (BA 44/45). These findings indicate functional differentiation of the fronto-temporal systems for processing spoken language, with left middle temporal gyrus (MTG) and superior temporal gyrus (STG) involved in mapping sounds to meaning, bilateral ventral inferior frontal gyrus (IFG) engaged in less constrained early competition processing, and bilateral dorsal IFG engaged in later, more fine-grained selection processes.

Keywords: spoken word recognition, lexical competition, lexical selection, inferior frontal gyrus, cohort model

Introduction

Cognitive models developed over the last several decades (e.g. Marslen-Wilson and Welsh 1978; Elman and McClelland 1986; Marslen-Wilson 1987; Norris 1994; Gaskell and Marslen-Wilson 1997) have put forward a range of detailed proposals about the mechanisms of spoken word recognition, proposing a set of fine-grained representations, processes, and structures. One of the key claims is that speech sounds are mapped onto meaning by an optimally efficient language processing system (e.g. Marslen-Wilson and Tyler 1981; Marslen-Wilson 1984; Norris and McQueen 2008). These models do not, however, have well-developed equivalents in the neural domain. This study aims to investigate the neural substrates of the human language processing system by relating neural activity to cognitive claims and behavioral data.

The term "optimal efficiency" emerged in this context as a predicted property of the "Cohort Model" of spoken word recognition, which was designed to explain the extreme earliness

of word recognition and its sensitivity to contextual constraints (Marslen-Wilson 1975; Marslen-Wilson and Welsh 1978). This model, and its later variants (Marslen-Wilson 1987; Gaskell and Marslen-Wilson 1997), assumed a fully parallelized recognition process (cf., Morton 1969; Fahlman 1979), where all word candidates that initially fit the accumulating speech input (the word-initial "cohort") are continuously assessed, and drop out of contention as mismatches emerge. Whether this model is viewed in virtual terms (as a property of a massively parallel neural network) or as reflecting the action of multiple independent "word detectors," it makes the prediction that a spoken word will be recognized as soon as the information becomes available in the speech stream that differentiates it from its competitors. This is optimally efficient in the sense that it makes maximally effective use of incoming sensory information to guide dynamic perceptual decisions.

This prediction of the cohort model was directly tested in a behavioral experiment, which used a nonword detection task to tap into the timing of lexical processing in spoken sequences (Marslen-Wilson 1984). This study showed that spoken nonwords could be recognized as nonwords as soon as the spoken sequence deviated from a real word, at the so-called "nonword point." This is the point in the speech sequence at which a potentially meaningful sequence becomes a nonword, and it varied in this study from the second to the fifth phoneme in the sequence. Average reaction times (RTs) to make a nonword decision were strikingly constant at around 450 ms relative to the nonword point, as reflected in the high correlation ($r = 0.72$) between item RTs and the duration from sequence onset to the nonword point. In contrast, the correlation between RTs and the duration of the whole sequence was not significant.

The stability of the nonword point effect across early and late divergence points indicates that a real-time analysis, relating the incoming speech input to possible words in the language sharing the same initial sequence, starts as soon as some minimum amount of information (e.g. one phoneme) is available from the speech input. The decision to reject the sequence as a nonword can therefore begin to be made as soon as there are no word candidates that still match the incoming sensory input. As information becomes available in the speech signal, it is used to guide perceptual choice between different word candidates—with nonword detection being one end-point of this process. However, the neural underpinnings of these processes, and of the nonword point effect in particular, remain unclear. The aim of the present study is to investigate the neural basis of this optimally efficient process by exploiting the nonword point effect.

The online efficiency of the language system does not imply that the recognition of a spoken word is simply a one-to-one direct mapping from the sound of the word onto a stored representation of that word's meaning. Instead, it involves the continuous activation of multiple competing word candidates and corresponding selection and decision processes (e.g. Marslen-Wilson and Welsh 1978; McClelland and Elman 1986; Marslen-Wilson 1987; Zwitserlood 1989; Norris 1994; Gaskell and Marslen-Wilson 1997, 2002; Allopenna et al. 1998). In the formulation adopted by the Cohort model (Marslen-Wilson and Welsh 1978; Marslen-Wilson 1987), there are 2 related processes involved in spoken word recognition: The activation of multiple cohort candidates, and the subsequent evaluation and rejection of inappropriate candidates.

Word-initial speech sounds (e.g. [æl] of "alligator") activate an initial cohort of simultaneously active word candidates (e.g. "alcohol," "albatross," "alligator"), which share the same initial sound sequence [æl]. This activation is claimed to be an autonomous process, exclusively driven by bottom-up sensory inputs. It is also necessarily accompanied by a resolution process as the multiple cohort candidates compete for selection and recognition. When the number of cohort candidates (cohort size) is larger, the competition among them becomes stronger (Marslen-Wilson and Welsh 1978; Marslen-Wilson 1990; Tyler et al. 2000). Once word candidates are activated in the initial cohort, they continue to be evaluated against the incoming sensory input, and the activation level of words that mismatch gradually declines (Tyler 1984; Marslen-Wilson 1987). The rejection of inappropriate candidates can be seen as a process of selection, affected by both bottom-up sensory inputs and top-down factors such as contextual constraints (Tyler and Wessels 1983; Tyler 1984) and lexical semantics (Tyler et al. 2000). Competition and selection are related but potentially separable processes, as suggested by research in related domains such as speech production (e.g. Mahon et al. 2007).

The brain mechanisms involved in competition and selection have been extensively investigated in previous neuroimaging studies, with left inferior frontal gyrus (LIFG) being identified as a critical region for the general function of selection (e.g. Thompson-Schill et al. 1997, 2005; Moss et al. 2005; Rodd et al. 2005, 2012; Grindrod et al. 2008; Schnur et al. 2009). In many of these studies, LIFG activity is accompanied by activation in the right (R) hemisphere homologous site, although right inferior frontal gyrus (RIFG) activity is typically weaker (e.g. Thompson-Schill et al. 1997; Badre and Wagner 2004; Rodd et al. 2005; Bilenko et al. 2008; Zhuang et al. 2011). The role of RIFG remains underspecified, leaving open whether it is as strongly involved in competition and selection functions as the LIFG—as claimed by Bozic et al. (2010)—or simply plays a complementary role in supporting the LIFG.

The activation of these competition and selection effects varies within bilateral IFG across different tasks, stimuli, and types of competition (e.g. phonological, semantic, and syntactic). Effects are seen in left Brodmann area (L BA) 44, for example, for word generation and picture classification tasks and L BA 44/45 and R BA 44 for a word comparison task (Thompson-Schill et al. 1997, semantic competition); in L BA 47/45/44 for a picture naming task (Moss et al. 2005, semantic competition); and in bilateral BA 44/45/47 for working memory retrieval (Badre and Wagner 2004). Within the domain of spoken word recognition, different activation patterns have been reported in inferior frontal cortex; for example, bilateral BA 45/47 (Bozic et al. 2010, gap detection with no gaps on test trials), bilateral BA 45/44 (Bilenko et al. 2008, lexical decision), L BA 45/47 (Grindrod et al. 2008, lexical decision), and L BA 45/47 and R BA 47 (Zhuang et al. 2011, lexical decision). One notable finding is that L BA 45 is the most commonly activated subregion of IFG across different tasks and stimuli, especially for spoken word recognition.

A further issue is how competition and selection, in so far as they are cognitively distinct processes, are underpinned by frontal involvement within the dynamic neural systems supporting spoken language processing. Previous neuroimaging research has rarely investigated the relationship between competition and selection, and the 2 processes are typically not separated from each other in the experimental manipulations used (e.g. Thompson-Schill et al. 1997). An exception is the recent study by Grindrod et al. (2008), which manipulated different semantic competition and selection conditions in spoken word recognition, using an implicit priming task. These authors found only selection effects in LIFG, and no effect of competition, though this may be because the competition effect is relatively weaker, and difficult to detect with an indirect paradigm such as implicit priming. Using an explicit task (lexical decision), Zhuang et al. (2011) observed a competition effect in the initial cohort of spoken words with greater activation in LIFG (BA 45/47) and RIFG (BA 47) for increasing cohort size.

Experimental Considerations

Within the framework of the Cohort model, the present study was designed to address 2 specific issues involved in recognizing spoken words: What are the neural underpinnings of this optimally efficient language processing system and how are the processes of competition and selection, which are a central part of this system, instantiated in the brain? To address these questions, we designed a 2-part study, with a behavioral component run outside the scanner, and a functional magnetic resonance imaging (fMRI) component run on the same stimulus set, manipulating the nonword point and 2 continuous variables indexing cohort competition and selection processes. The purpose of the behavioral component was to establish unequivocally that the nonword point effect was elicited by the current stimuli, validating the basic cognitive claim about the dynamic functional properties of the recognition system. The goal of the fMRI component is to explore the architecture of the neural systems supporting these capacities.

Nonword sequences, modeled on the original Marslen-Wilson (1984) study, were constructed so that at sequence onset each stimulus could potentially be a real word. The primary manipulation was the position of the nonword point —whether it occurred early, middle, or late in the sequence. The nonword point was measured in 2 ways—either by the duration from sequence onset to nonword point or by the amount of phonological information, as measured by the number of phonemes heard at early, middle, and late nonword points. Sequences in the early nonword point condition became nonwords at the beginning of the second or third phoneme (e.g. at the [v] in "kvint", [kvint], or the [au] in "smaud", [smaud]). To maintain comparability with the

original Marslen-Wilson study, we included a distinction between sequences that are phonotactically illegal, such as "kvint", where [kv] cannot appear word initially in English, and sequences such as "smaud", where the initial sequence [sm] is phonotactically legal. However, this contrast addresses issues outside the scope of this report and its results are not presented in detail here.

Sequences in the middle and late nonword point conditions (all phonotactically legal) became nonwords, respectively, at the consonant after the initial vowel (for example, at the [v] in "soivish", [soivi/]) or at the following consonant/vowel (for example, at the [d] in "trandal," [trændəl] or the [ei] in "skoonate," [sku:neit]). Behaviorally, we expected to elicit the pattern reported by Marslen-Wilson (1984), with a strong positive correlation between rejection latencies and "prenonword point duration," the duration from sequence onset up to nonword point. At the same time, the duration from nonword point until the end of the sequence ("post-nonword point duration") should be less effective in predicting the RTs for rejecting a nonword. Secondly, there should be a factorial effect of nonword point with slower rejection times for sequences with later-appearing nonword points.

We also expect specific neural reflexes of these nonword point effects. These predictions follow intrinsically from a sequential cohort process, where as more of a sequence is heard, the more extensive the match will be to the remaining members of the cohort. This means that sequences with later nonword points should generate stronger lexical-semantic activation in areas of the brain supporting the primary processes of lexical access - namely, the L temporal regions that mediate the link between incoming phonological information and underlying lexical representations (e.g. Binder et al. 1996, 1997; Dronkers et al. 2004; Indefrey and Cutler 2004). This predicts greater brain activation in L temporal cortex for the later nonword point sequences. The neural network most activated at later nonword points should be central to the process of mapping speech sounds onto meaning, since this is the network that will have the most sustained lexical access up to and including the point at which the nonword rejection decision is made.

The second aim of the fMRI component is to investigate the neural substrates of competition and selection processes in accessing spoken sequences by manipulating the size of the initial cohort and the cohort drop-out rate. Initial cohort size refers to the number of words in the language (as known to the listener) sharing the same initial phonemes, defined here as either the initial 2 consonants (CC) or the initial consonant and vowel (CV). When cohort size is larger, the competition among its members is potentially higher than when a cohort contains few candidates (Marslen-Wilson and Welsh 1978; Marslen-Wilson 1990; Tyler et al. 2000). We will investigate the neural underpinning of cohort competition by correlating neural activity with initial cohort size and predict greater activation in frontal cortex as cohort size increases (cf., Thompson-Schill et al. 1997; Moss et al. 2005; Schnur et al. 2009; Zhuang et al. 2011).

The selection process, in contrast, is represented by cohort drop-out rate, defined as the ratio of the terminal cohort size to the initial cohort size (both log-transformed). The terminal cohort size refers to the number of word candidates sharing the same phonemes up to and including the phoneme before the nonword point. The cohort drop-out rate measures the drop-out speed of word candidates from initial to terminal cohorts. As the drop-out rate becomes higher, more word candidates drop out of the cohort, reflecting a more intensive process of selection. This drop-out process involves automatically evaluating and selecting word candidates, so that candidate brain regions for this effect should be bilateral IFG (BA 44, 45, and 47), which has been claimed to be critical in general selection (e.g. Thompson-Schill et al. 1997, 2005; Moss et al. 2005; Rodd et al. 2005, 2012; Bilenko et al. 2008; Grindrod et al. 2008; January et al. 2008; Schnur et al. 2009; Bozic et al. 2010; Zhuang et al. 2011).

Part 1: The Behavioral Component

We first performed a behavioral experiment, building on previous findings (Marslen-Wilson 1984) that implicated an optimally efficient language processing system, to determine whether manipulations of nonword point provided an appropriate foundation for the planned imaging component.

Materials and Methods

Participants

Eighteen healthy volunteers (6 males and 12 females, aged 19–34) who were native British English speakers with normal hearing took part in this experiment. They all gave informed consent and were compensated for their time.

Design and Materials

The stimuli were made up of 360 nonwords with 360 real words from another study (Zhuang et al. 2011) acting as fillers. The real word fillers consisted of both concrete and abstract words from the CELEX database (Baayen et al. 1995), with 94 monosyllabic, 169 disyllabic, and 197 trisyllabic items. The 360 nonwords were each assigned to 1 of the 4 experimental conditions, varying in their nonword point (as described earlier): Early nonword point, phonotactically illegal (72 items), early nonword point, phonotactically legal (72 items), middle nonword point (72 items), and late nonword point (144 items). The overall lengths of nonwords varied from 1 to 3 syllables. There were equal numbers of monosyllabic, disyllabic, and trisyllabic nonwords in early and middle nonword point conditions. Because the late nonword point occurs at the beginning of the second syllable, there were no monosyllabic nonwords for the late nonword point conditions, which had equal numbers (72) of disyllabic and trisyllabic items. In total, there were 72 monosyllabic, 144 disyllabic, and 144 trisyllabic items.

For the 144 sequences with late nonword points, there were 2 additional manipulations, the initial cohort size and the cohort drop-out rate, as measures of cohort competition and selection processes. The number of stimuli (144) in the late nonword point condition made it possible to explore the effects of these variables using correlational methods.

The stimuli were recorded by a female native speaker of British English onto a digital audio tape recorder at a sampling rate of 44100 Hz, and then transferred to a computer, where they were downsampled using CoolEdit Software to a lower rate (22 050 Hz, 16 bit resolution, monochannel) for presentation with the experimental software. Each stimulus was placed in an individual speech file. The mean duration for the nonwords was 782 ms, with a range from 398 to 1183 ms (standard deviation [SD] = 159 ms). The mean duration of the real word fillers was 548 ms (SD = 115 ms).

Once each speech file had been created, the time to the nonword point was measured in milliseconds from sequence onset to the onset of the nonword point phoneme. In some sequences, it is difficult to identify the exact onset of this phoneme because of the overlap between successive phonemes in the waveform. In these cases, we first estimated the approximate middle point between 2 adjacent phonemes by identifying their waveform peaks, then adjusted the

Table 1

Description of the nonword stimuli with standard deviation in brackets

Condition	Description	Nonword point	Item number	Sequence duration (ms)	Pre-nonword point duration (ms)	Example
Early nonword point	Phonotactically illegal	CC	72	807 (178)	136 (71)	k<u>v</u>int
Early nonword point	Phonotactically legal	(C)CV	72	772 (150)	161 (69)	sm<u>au</u>d
Middle nonword point	Phonotactically legal	(C)CVC	72	724 (177)	293 (93)	soi<u>v</u>ish
Late nonword point	Phonotactically legal	(C)CVCC/ (C)CVC<u>V</u>	144	803 (135)	298 (78)	skoo<u>n</u>ate

Note: C = consonant, V = vowel; nonword points are underlined for each condition.

nonword point as necessary by listening to successive increments of the speech signal. The average duration to each of the nonword points is shown in Table 1. Duration to the nonword point was similar in the middle and late conditions (293 vs. 298 ms, Table 1), even though the late nonword point occurs on average one phoneme later. This is because speakers typically pronounce trisyllabic sequences at a faster rate than disyllabic sequences (Lehiste 1972; Klatt 1976).

Procedure

Participants were tested in a sound-attenuated room in groups of up to four. Participants heard the stimuli presented via headphones using DMDX experimental software (Forster and Forster 2003). Participants were asked to decide whether the stimulus they heard was a real word in English or not. They pressed the "yes" button of a response box when they heard a real word and the "no" button when they heard a nonword. They were asked to do this as quickly as possible. The time out was set at 3 s, and the intertrial interval was 1.5 s.

There was a practice session of 24 items. There were 4 experimental sessions, each beginning with 4 lead-in items. Items from each condition were evenly distributed across the different sessions, and the order of presentation was pseudorandomized. There were not more than 3 adjacent items from the same condition and not more than 4 adjacent real words or nonwords. The order of presentation of the experimental sessions was varied between participants.

Results

The data from 1 participant were excluded due to very slow response times (mean = 1007 ms; group mean = 770 ms) and 1 item ("wike") was excluded due to its high error-rate (64.7%). This left a total of 17 participants and 359 items. Only correct responses (96.9%) were included in the RT analyses. RTs were inverse-transformed to reduce the effects of outliers (Ratcliff 1993; Ulrich and Miller 1994), then analyzed across items ($F2$). Subject analyses ($F1$) were omitted as extraneous variables cannot be taken as covariates. Error analyses are not reported as the overall error rate was very low (2.4%), and not more than 3.1% in any of the 3 conditions.

The primary predictions concerned the nonword point effect, which we analyzed using both correlational and factorial techniques. The correlation analysis showed a significant correlation between RTs and pre-nonword point duration, $r = 0.61$, $P < 0.001$ (Fig. 1A). As the duration from sequence onset to nonword point increases, time to reject the sequence as a nonword becomes longer, consistent with claims for the continuous activation of lexical information as the sequence unfolds up to the nonword point (and beyond). In contrast, the correlation between RTs and the post-nonword point duration was much weaker and failed to reach significance

($r = -0.096$, $P = 0.07$). These results confirm the findings in Marslen-Wilson (1984), where response times were strongly and linearly dominated by pre-nonword point duration.

We then performed a factorial analysis on the 3 nonword point conditions, collapsing across the phonotactically legal and illegal early conditions. In a 1-way analysis of variance, RTs were taken as a dependent variable with the 3 nonword point conditions as factors. There was a significant effect of nonword point, $F2_{2,356} = 21.35$, $P < 0.001$, with longer response times to sequences with later-appearing nonword points. In post hoc tests using least significant difference (LSD), sequences in the early nonword point condition (mean RT = 736 ms) were responded to more quickly than those in the middle (mean RT = 762 ms; $P = 0.015$) and late (mean RT = 793 ms, $P < 0.001$) conditions, and sequences in the middle condition were recognized faster than those in the late condition, $P = 0.005$. Figure 1B shows the tendency for a linear relationship across these 3 nonword point conditions in mean RTs. These results are consistent with the correlational analyses and confirm that the nonword point is the most important factor in determining the lexical decision response.

We also looked for potential behavioral effects of variations in cohort competition and selection, testing the 144 sequences with late nonword points for correlations between RTs and initial cohort size and between RTs and cohort drop-out rate. For cohort competition, 3 extraneous variables (pre- and post-nonword point durations, and the summed initial cohort frequency) were partialled out as covariates. Since word candidates in the initial cohort vary in frequency, the activation level of each of these candidates varies as well. To reduce the potential influence of cohort activation on competition, we partialled out the summed initial cohort frequency, assuming that each cohort candidate was equally weighted in frequency with the same activation level, and that cohort size was an effective measure of the degree of competition between candidates within a cohort. There was a significant positive correlation between RTs and initial cohort size, $r = 0.24$, $P = 0.004$. Response latencies to reject a nonword were slower as the size of the initial cohort increased, possibly reflecting increased competition.

To evaluate possible behavioral effects of cohort selection, we partialled out 4 extraneous variables: Pre- and post-nonword point durations, the summed initial cohort frequency, and the initial cohort size. The reason for covarying out initial cohort size and frequency was to minimize the influence of cohort competition on selection, given that cohort competition at an earlier stage in time could potentially affect a later stage selection process, but not vice versa. Since the correlation between the initial cohort size and cohort drop-out rate is relatively low ($r = 0.31$), it is theoretically plausible to separate the 2 effects. The correlation between RTs and cohort drop-out rate was marginally significant, $r = 0.15$, $P = 0.087$. As the drop-out rate increases, with larger number of word candidates being evaluated and discarded, RTs also tend to increase.

Discussion

The strong nonword point effects in this study, with a revised stimulus set and a different task, confirm that the nonword point is a critical factor in determining the rejection latencies of nonwords. When the nonword point
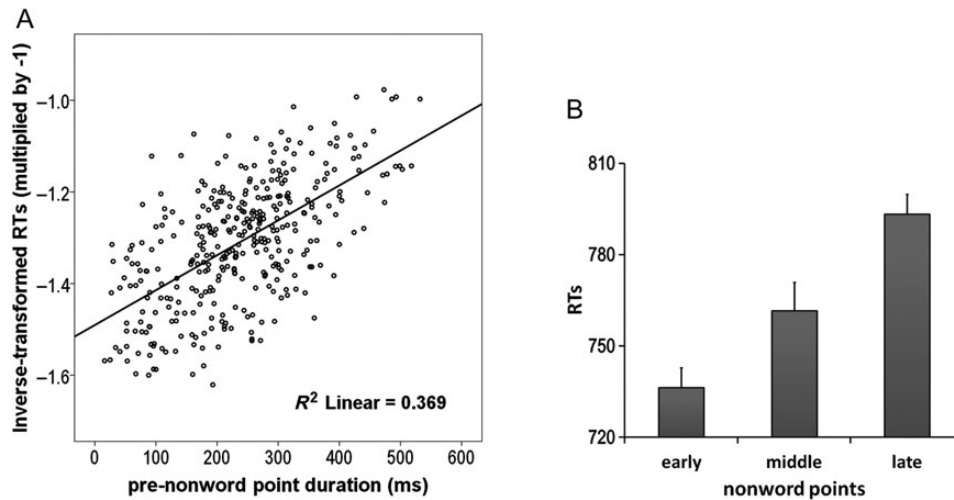
Figure 1. Results of behavioral study. (*A*) Item-wise pre-nonword point duration plotted against RT. (*B*) Mean lexical decision RTs at early, middle, and late nonword points.

occurs later in a spoken sequence, the sequence remains a potential real word for longer, and response latency is slower. The results of the correlational analysis replicate the general findings of Marslen-Wilson (1984), although the correlation was lower than in the original study ($r = 0.61$ rather than 0.72). This may well reflect task differences. The Marslen-Wilson (1984) study used a nonword detection task, where participants only responded to the nonwords and gave no response to the real words. In these circumstances, where "yes" decisions are not being made to real words (which requires all of the sequence to be heard before the response can be made), responses can be more tightly tuned to the within-sequence nonword point.

We also saw evidence for the influence of lexical competition and selection on word recognition, as reflected in the initial cohort size effect and cohort drop-out rate effect. Initial cohort size correlated positively with RTs, when extraneous variables were partialled out. In a large cohort, with a larger number of candidates, it may take longer to resolve cohort competition during the initial stages of lexical access. The cohort drop-out rate also seems to influence the recognition of spoken words, with slower response times when more word candidates are being discarded from the cohort, involving higher demand on selection processes.

Part 2: The fMRI Component

Based on the behavioral foundation provided by Experiment 1, we performed an fMRI study using the same task and stimuli, to investigate the brain mechanisms supporting this optimally efficient mapping from sound to meaning, and the neural underpinnings of the competition and selection functions that play key roles in this process.

Materials and Methods

Participants

Fourteen healthy control volunteers (7 males and 7 females, aged 19–33 years) took part in the fMRI study. All were native English speakers with normal hearing and were right-handed as assessed by the Edinburgh Handedness Inventory (Oldfield 1971). They all gave informed consent and were compensated for their time. The study

received approval from the Peterborough and Fenland Local Research Ethics Committee, UK.

Design and Materials

The materials (360 nonwords) were the same as those used in Experiment 1 with the addition of 80 null events (silence) to provide a baseline condition. However, the design was slightly different from that in the behavioral experiment. Since the main purpose of this study was to use the nonword point effect to investigate the neural interface that maps sounds onto meaning, only the linguistic (phonemic) changes from early to middle nonword point and from middle to late nonword point were of theoretical interest. Any duration differences among these conditions were treated as extraneous variables in the imaging analyses. These include the variations in pre- and post-nonword point durations, which may produce activation in the same temporal regions as the hypothesized nonword point effect.

Two types of parametric modulation designs were used to explore the nonword point effect and the cohort competition and selection effects. The analysis of the nonword point effect was analogous to the factorial analysis of the same effect in the behavioral data. All 360 nonwords in the 4 experimental conditions were included. Sequence duration variations across items, including pre- and post-nonword point durations, were partialled out as extraneous variables, since they were not matched across conditions.

In contrast, only nonwords with late nonword points (144 items) were used to explore cohort competition and selection processes. Sequences with early or middle nonword points become nonwords so early that their initial and terminal cohorts are almost the same. The parametric modulation analyses on the 3 cohort variables (initial cohort size, summed initial cohort frequency, and cohort drop-out rate) were similar to the partial correlational analyses in the behavioral data. For example, to test the cohort competition effect, the initial cohort size would be correlated with neural activity with the initial cohort frequency as a covariate, in which case the members within each cohort were treated equally (frequency weighted).

Procedure

The previously recorded stimuli underwent a further process of pre-emphasis prior to presentation (http://www.mrc-cbu.cam.ac.uk/~rhodri/headphonesim.htm), in order to optimize auditory quality in the acoustically challenging scanner environment. Stimuli were presented using CAST experimental software (http://www.mrc-cbu.cam.ac.uk/~maarten/CAST.htm) and were delivered to participants via Etymotic ER3 insert earphones.

Participants were instructed to respond to each stimulus by pressing a response key with their index finger for real words and middle

finger for nonwords, and to make no response to the baseline (silence) items. Items were divided into 4 sessions with items from each condition evenly distributed across the sessions. Within each session, the order of presentation of real words, nonwords, and silence was pseudorandomized, such that not more than 4 real words or nonwords followed one another. This pseudorandomization was also applied to the 4 experimental conditions. Session order was counterbalanced across participants. Each session consisted of 200 experimental trials with 5 lead-in dummy scans for MR signal stabilization and 2 dummy scans at the end. Each session lasted 12 min, and participants had a brief rest between sessions. Before the first session, there was a short practice session of 12 items to familiarize participants with the procedure inside the scanner. During the experiment, both RTs and errors were recorded.

MRI Acquisition and Imaging Analysis

Scanning was performed on a 3-T Magnetom Trio (Siemens, Munich, Germany) at the MRC Cognition and Brain Sciences Unit, Cambridge, UK, using a gradient-echo echo-planar imaging (EPI) sequence (repetition time = 3400 ms, acquisition time = 2000 ms, echo time = 30 ms, flip angle 78°, resolution $3 \times 3 \times 3.75$ mm, matrix size 64×64, field of view 192×192 mm, 32 oblique slices away from the eyes, 3 mm thick, 25% of slice gap) with head coils, 2232-Hz bandwidth and spin-echo-guided reconstruction. We acquired T1-weighted MPRAGE scans for anatomical localization. We used a fast sparse imaging protocol (Hall et al. 1999) in which speech sounds were presented in the 1.4 s of silence between scans. There was a silent gap of 100 ms between the end of each scan and the onset of the subsequent stimulus, minimizing the influence of preceding scanning noise on the perception of sequences, especially their onsets. The time between successive stimuli was jittered to increase the chance of sampling the peak of hemodynamic response.

Preprocessing and statistical analysis were carried out in SPM5 (Wellcome Institute of Cognitive Neurology, London, UK; www.fil.ion. ucl.ac.uk), under MATLAB (Mathworks, Inc., Sherborn, MA, USA). All EPI images were realigned to the first EPI image (excluding the 5 initial lead-in images) to correct for head motion, and then spatially normalized to a standard MNI (Montreal Neurological Institute) EPI template, using a cut-off of 25 mm for the discrete cosine transform functions. Statistical modeling was done in the context of the general linear model (Friston et al. 1995) as implemented in SPM5, using an 8-mm full-width half-maximal Gaussian smoothing kernel.

In the fixed effect analysis for each participant, a parametric modulation design (Buchel et al. 1996; Henson 2004) was used to model the experimental conditions. Two slightly different modulation methods were applied in this study according to the research aims and experimental design. To investigate the nonword point effect, the first analysis was performed by taking the 4 nonword point experimental conditions as modulators with binary values (0, 1). In each of the 4 testing sessions, the design matrix consisted of 15 columns of variables: Nonwords, pre- and post-nonword point durations, 4 experimental conditions (early nonword point/phonotactically illegal, early nonword point/phonotactically legal, middle nonword point, and late nonword point), real words, null events (baseline), and 6 movement parameters. Among these variables, there were 3 independent events: Nonwords, real words, and null events. The nonword event was modulated by 6 parametric modulators—pre- and post-nonword point duration and the 4 experimental conditions. For each column representing a single condition, every nonword item belonging to this experimental condition was labeled as "1", and the other nonword items were labeled as "0". For example, the item "kvint" was labeled as "1" in the modulator column for the early nonword point and phonotactically illegal condition, and items from the other 3 experimental conditions (e.g. "smaud", "soivish", "skoonate") were labeled as "0" in the same column. At the same time, the item "kvint" was labeled as "0" for the modulator columns corresponding to the other 3 experimental conditions. Each of the 6 modulators was orthogonalized relative to the other 5 modulators, so that any shared variance among these modulators was removed, and any difference among these 4 experimental conditions in duration was partialled out.

The second analysis was performed to investigate cohort competition and selection effects. The nonwords with late nonword points (144 items) were taken as an independent event, which was modulated by 5 modulators in the following order: Pre- and post-nonword point duration, the summed frequency of the initial cohort members, the initial cohort size, and the cohort drop-out rate. In accordance with the behavioral data analyses, the same 3 extraneous variables (the first 3 modulators here) were partialled out for the cohort competition effect (initial cohort size), and the same 4 extraneous variables (the first 4 modulators) were partialled out for the cohort selection effect (the cohort drop-out rate). To do this, we treated the 5 modulators sequentially, from left to right, so that any shared variance between any 2 modulators was allocated to the earlier modulator in order. The same design matrix also included 6 movement parameters and 3 extra independent events: Nonwords in the other 3 conditions, real words, and null events. Error items were removed from the nonword events in both analyses.

Trials were modeled using a canonical hemodynamic response function (HRF), and the onset of each stimulus was taken as the onset of the trial in the SPM analysis model. The data for each participant were analyzed using a fixed effect model and then combined into a group random effect analysis. Activations were thresholded at $P < 0.005$, uncorrected, at the voxel level, and significant clusters were reported only when they survived $P < 0.05$, cluster-level corrected for multiple comparisons unless otherwise stated. SPM coordinates were reported in MNI space. Regions were identified by using the AAL atlas (Tzourio-Mazoyer et al. 2002) and Brodmann templates as implemented in MRIcron (www.cabiatl.com/mricro/mricron). As this study aimed to explore activations within the neural language processing system, a mask covering the neural regions typically considered to encompass the fronto-temporo-parietal language network was consistently used for all imaging results reported in this study (e.g. Boatman 2004; Dronkers et al. 2004; Indefrey and Cutler 2004; Scott and Wise 2004; Hickok and Poeppel 2007; Tyler and Marslen-Wilson 2008). The mask was created using the WFU PickAtlas (Maldjian et al. 2003, 2004) and comprised the following regions defined by the AAL atlas: Bilateral IFG (BA 44, 45, 47), orbitofrontal cortex (BA 47, 11), Rolandic operculum, anterior cingulate gyrus, insula, superior temporal gyrus, middle temporal gyrus, angular gyrus, supramarginal gyrus, and inferior parietal lobule.

Results

Behavioral Results

Two items ("thooton", "thel") were removed from the analysis due to high error rates (over 90%), leaving 358 nonword items. Where participants made error responses (4.7%), the data were excluded from the analyses. The RTs were inverse-transformed to reduce the influence of outliers (Ratcliff 1993; Ulrich and Miller 1994). Only item analyses ($F2$) were performed. Error analyses are not reported as the overall error rate was low (4.7%), and not more than 5.5% in any condition. RTs were generally slower (1045 ms overall) than in the behavioral study (794 ms), reflecting the less favorable auditory environment in the scanner.

Nevertheless, in both correlational and factorial analyses, the data again reveal significant nonword point effects. In a Pearson correlation analysis, pre-nonword point duration was found to be significantly correlated with RTs, $r = 0.38$, $P < 0.001$, with slower RTs to later occurring nonword points. A significant correlation was also found with post-nonword point duration, $r = 0.20$, $P < 0.001$, most likely reflecting the slower response times in the scanner. However, this was a significantly weaker predictor of RT than pre-nonword point duration ($t(355) = 1.98$, $P < 0.05$). With post-nonword point duration partialled out as an extraneous variable, the
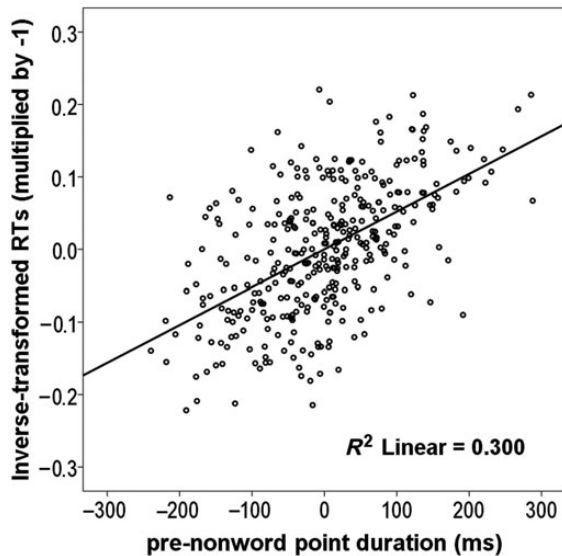
Figure 2. Behavioral data in the scanner. Pre-nonword point duration plotted against RT (with post-nonword point duration covaried out).

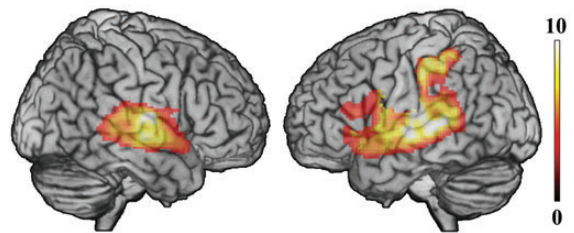| Regions	BA	Cluster-level		Voxel-level		MNI coordinates		
		$P_{corrected}$	Extent	$P_{corrected}$	Z	x	y	z
L STG, MTG, IFG, IPL, SMG, Rolandic operculum, insula	41, 42, 21, 22, 38, 40, 44, 45, 47	0.000	8591	0.000	6.15	−56	−18	12
				0.000	5.82	−58	−36	6
				0.000	5.79	−64	−28	2
R STG, MTG, Rolandic operculum, insula	41, 42, 21, 22	0.000	3530	0.001	5.51	64	−8	10
				0.009	5.06	66	−14	−4
				0.038	4.77	58	−4	−6
Bilateral anterior cingulate	24	0.000	583	0.002	5.34	6	24	28
				0.005	5.19	−4	24	30
				0.051	4.69	−10	18	30

Figure 3. Significant activation for the contrast of nonwords minus baseline (silence) at a threshold of $P < 0.005$, voxel-level uncorrected, and $P < 0.05$, cluster-level corrected. Color scale indicates t-value of contrast.

correlational nonword point effect for pre-nonword point duration increased to $r = 0.55$, $P < 0.001$ (Fig. 2), comparable with the behavioral results seen outside the scanner.

Secondly, a factorial analysis of covariance was carried out on the 3 nonword point conditions (collapsing as before over phonotactic early conditions), with post-nonword point duration partialled out as a covariate. There was a significant effect of nonword point, $F2_{2,354} = 6.54$, $P = 0.002$, with longer response times to sequences with later nonword points. In post hoc comparisons using the LSD test, sequences with the early nonword points (mean RTs = 993 ms) were rejected faster (mean difference = 36 ms) than those with late nonword points (RT = 1029 ms), $P = 0.002$, and sequences with middle nonword points (RT = 995 ms) were recognized faster (34 ms) than those with late nonword points, $P = 0.010$. There was no significant difference between sequences with early and middle nonword points, $P > 0.1$.

These analyses confirm that the participants in the fMRI component exhibited a functionally equivalent pattern of responses to those seen earlier, with the speech input being continuously evaluated against stored representations of possible words in the language. The analysis of the imaging data focuses on the neural systems supporting these capacities.

Imaging Results
The first step of the imaging analyses was to test the effectiveness of the experimental task and stimuli in eliciting brain activation in the spoken language processing network. To this end, we compared the activations resulting from all nonwords against the null events (silent baseline). This analysis produced significant activation in bilateral STG (BA 41, 42, 21, 22, and 38), MTG (BA 21 and 22), anterior cingulate (BA 24), L IFG (BA 44, 45, 47), inferior parietal lobule (BA 40), supramarginal gyrus (40), Rolandic operculum, and insula (Table 2 and Fig. 3). These regions are typically activated in neuroimaging studies of spoken language (e.g. Price et al. 1996; Binder et al. 2000; Davis and Johnsrude 2003; Tyler et al. 2005).

To determine the neural substrate mediating sound-meaning mapping, we used a nonword point measure. This

was calculated using the parametric modulation design matrix described above where the 4 nonword point conditions were taken as modulators. In the same design matrix, the 2 extraneous duration variables were also included as modulators. Pre- and post-nonword point durations were partialled out, as discussed above, in order to separate the neural effects of irrelevant variations in duration from the effects of interest related to the nonword points themselves.

A correlational analysis on the nonword point conditions showed a positive nonword point effect in L anterior and middle MTG (BA 21, 22), extending to STG (BA 22), with the peak in L MTG (BA 22, −60 −10 −8; Table 3 and Fig. 4A). Later nonword points generated stronger neural activity in these brain regions. To explore further the role of each condition in the nonword point effect, we selected the activated cluster shown in Figure 4A as a region of interest, within which we examined the activation of the 4 nonword point conditions using MarsBaR (http://marsbar.sourceforge.net/). Figure 4B shows a linear increase in activation for later nonword points (with the phonotactically legal and illegal conditions—which did not differ, $t(13) < 1$—treated as a single condition). Increased activity was observed for middle nonword points relative to early nonword points, $t(13) = 2.32$, $P < 0.05$ ($P = 0.037$) and for late nonword points relative to middle nonword points, $t(13) = 4.79$, $P < 0.001$. This result is consistent with our key prediction that processing demands will increase in left temporal regions for later nonword points, reflecting a more extensive match between the accumulating speech input and stored lexical representations.

Turning to issues of cohort competition and selection processes in the analysis of nonwords with late nonword points (144 items), correlational analyses were performed on initial cohort size, and cohort drop-out rate, respectively. There was a positive correlation between increasing cohort size and neural activity, mainly in bilateral BA 47 (pars orbitalis and

Table 3

Areas of activity for the effects of nonword point, cohort competition, and selection

Regions	BA	Cluster-level		Voxel-level		MNI coordinates		
		$P_{corrected}$	Extent	$P_{corrected}$	Z	x	y	z
Nonword point effect								
L MTG, STG	21, 22	0.000	1031	0.023	4.56	−60	−10	−8
				0.398	3.7	−50	12	−20
				0.727	3.39	−52	−18	−14
Cohort competition effect								
R pars orbitalis, orbitofrontal cortex	47	0.044	258	0.833	3.53	34	46	−12
				1.000	2.64	50	36	−12
L pars orbitalis, orbitofrontal cortex	47, 45	0.035	274	0.984	3.19	−50	20	−10
				0.991	3.13	−50	40	−14
				0.993	3.10	−42	54	−2
Cohort competition effect[a]								
L pars orbitalis, orbitofrontal cortex	47, 45	0.015	542	0.578	3.80	−22	48	−16
				0.984	3.19	−50	20	−10
				0.991	3.13	−50	40	−14
R pars orbitalis, orbitofrontal cortex[b]	47	0.057	398	0.833	3.53	34	46	−12
				1.000	2.64	50	36	−12
Cohort selection effect								
L pars opercularis, pars triangularis, Rolandic operculum, insula	44, 45	0.046	283	0.273	4.09	−34	2	18
				0.640	3.70	−32	−8	18
				0.828	3.49	−44	10	18
Cohort selection effect[a]								
L pars opercularis, pars triangularis, Rolandic operculum, insula	44, 45	0.033	511	0.273	4.09	−34	2	18
				0.640	3.70	−32	−8	18
				0.828	3.49	−44	10	18
R pars triangularis	45	0.039	488	0.345	4.00	44	34	16
				0.784	3.55	52	36	14
				0.981	3.15	46	30	6

[a]Activation at a lower threshold of $P < 0.01$, voxel-level uncorrected.
[b]Activation at a significant threshold of $P < 0.06$, cluster-level corrected.
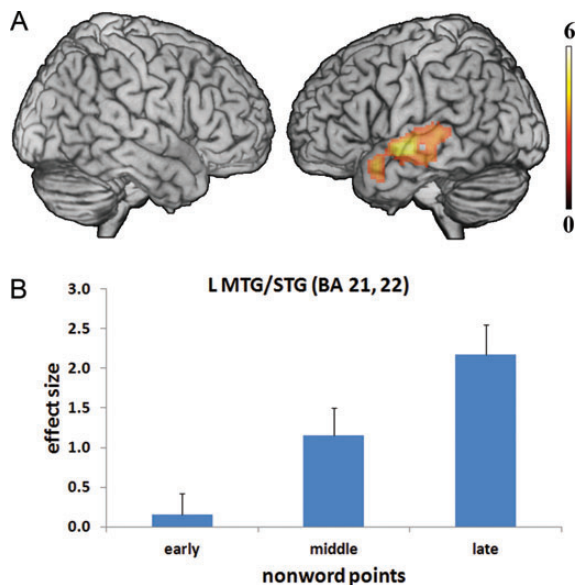
Figure 4. (A) Increasing activation for later nonword points at a threshold of $P < 0.005$, voxel-level uncorrected, and $P < 0.05$, cluster-level corrected. (B) The mean activation value of the significant cluster in (A) for each nonword point condition.

orbitofrontal cortex), extending into L BA 45 (pars triangularis; Table 3 and Fig. 5A). The larger the initial cohort, the greater the activation in these frontal regions, reflecting increased competition. To check whether any other regions were also involved in the competition processing, we lowered the voxel threshold to $P < 0.01$, uncorrected, and $P < 0.05$, cluster-level corrected. We again observed significant
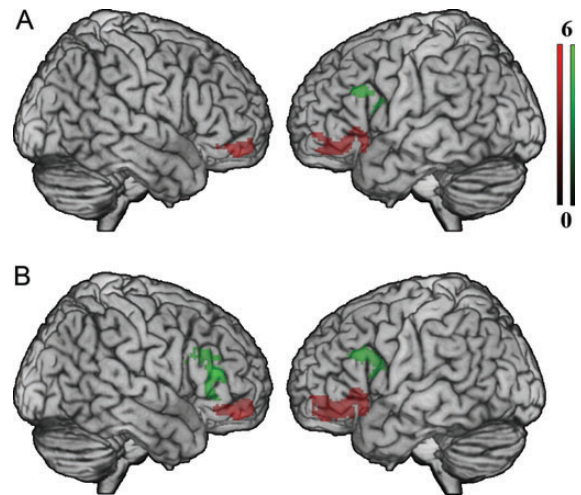
Figure 5. (A) Significant activation for increasing cohort competition (red) and increasing cohort selection (green) at a threshold of $P < 0.005$, voxel-level uncorrected, and $P < 0.05$, cluster-level corrected. (B) The cohort competition and selection effects rendered at a lower threshold of $P < 0.01$, voxel-level uncorrected, and $P < 0.06$, cluster-level corrected.

competition effects in bilateral BA 47 and L BA 45 (Table 3 and Fig. 5B), but no significant activation was seen in temporal, parietal, or occipital cortices.

For cohort selection, there was a positive correlation between increasing cohort drop-out rate and neural activity, focused in L BA 44 (pars opercularis), and extending to L BA 45 (pars triangularis), Rolandic operculum, and insula (Table 3 and Fig. 5A). Greater neural resources were recruited for higher selection demands - namely, when more word candidates drop out of the cohort as the sequence moves from initial to terminal cohort. Previous studies (e.g. Bozic et al. 2010; Zhuang et al. 2011) have suggested that processes of competition and selection involve a bilateral frontal system, although the involvement of RIFG may be weaker than that of the LIFG. When voxel threshold was lowered to $P < 0.01$, uncorrected, and $P < 0.05$, cluster-level corrected, we found that the cohort drop-out rate effect did include both LIFG and RIFG, with a significant cluster in R BA 45 (pars triangularis), (Table 3 and Fig. 5B). Consistent with the cohort competition analyses, we did not find significant activation in temporal, parietal, or occipital cortices related to cohort selection at either the default voxel threshold of $P < 0.005$, or at the lower threshold of $P < 0.01$, uncorrected, and $P < 0.05$, cluster-level corrected.

Discussion

This study investigates the brain mechanisms involved in spoken language processing in the context of the cognitive claims of the classic Cohort approach. The results demonstrate the validity and advantages of an approach that combines neuroimaging techniques with well-established cognitive models of the relevant domain. Overall, we found that a bilateral network of frontal and temporal regions is involved in spoken word recognition, and that activity within this network is modulated by different cognitive components of the word recognition process. There was a significant nonword point effect, focused in the LH and involving

anterior and middle MTG/STG (BA 21, 22) with greater activation for sequences with later nonword points. A cohort competition effect was found in bilateral ventral inferior frontal regions (BA 47/45), and a cohort selection effect bilaterally in more dorsal IFG (L BA 44/45, R BA 45), showing greater neural activity in these frontal regions for increasing cohort competition and selection.

The results of the behavioral component of the study reaffirm, for a new stimulus set and for a different task, the nonword point effect first reported by Marslen-Wilson (1984), and provide new evidence for optimal efficiency in the spoken word recognition process (Marslen-Wilson and Welsh 1978; Marslen-Wilson and Tyler 1981; Norris and McQueen 2008). The behavioral nonword point effect taps directly into the dynamic functioning of this system. Once a minimal amount of information (the initial phonemes of a sequence) is heard, the system starts analyzing the available information immediately, activating a cohort of word candidates sharing these initial phonemes. As the sequence unfolds, the system responds dynamically to the evolving sensory input. Only words which still match the incoming sensory inputs remain in the cohort, while others drop out. Relative to the point where no word candidate matches the sensory input, the sequence begins to be rejected as a nonword. It is the pre-nonword point duration, rather than the post-nonword point duration, that predicts latencies to reject sequences as nonwords.

The fMRI component of the study made it possible to take these powerful behavioral phenomena and to begin to map out the specific neural events associated with this optimally efficient mapping from sound to meaning. The L anterior and middle MTG/STG (BA 21, 22) activation associated with the nonword point is consistent with previous findings that these regions are involved in accessing stored lexical representations (e.g. Scott et al. 2000; Narain et al. 2003; Humphries et al. 2006; Spitsyna et al. 2006). This effect suggests that the L anterior and middle MTG/STG is involved in the interface between the mapping of speech sounds and meaning, since the recognition of a sequence (either as a nonword or real word) is completed at the nonword point (or the recognition point for a real word). Lexical processing is sustained as long as word candidates remain in the cohort. Sequences with later nonword points elicit more sustained lexical processing and therefore greater activity in L anterior and middle MTG/STG.

Cognitive models have claimed that spoken word recognition involves both activation of word candidates and processes that select among these competing alternatives (McClelland and Elman 1986; Gaskell and Marslen-Wilson 1997). The present study provides new insights into how the relevant competition and selection functions are instantiated in the brain and provides evidence for their neural separability—consistent with the relatively low correlation ($r = 0.31$) between the 2 markers of these functions (initial cohort size and cohort drop-out rate).

A cohort competition effect, reflecting variations in initial cohort size, was found mainly in bilateral BA 47 (pars orbitalis and orbital frontal cortex), extending into L BA 45 (pars triangularis), with increasing neural activity for larger cohorts with more word candidates. This competition effect is likely to reflect the relative indeterminacy of the cohort candidate set at the initial activation stage, where there is insufficient sensory information to direct selection to specific subsets of word candidates, and any member of the initial cohort is a potential recognition candidate.

In contrast, the cohort selection effect, reflecting variations in cohort drop-out rate, taps into later stages in the selection process, where we hypothesize that the recognition process is moving from the activation of an initial, coarsely defined set of potential word candidates to the specific analysis of the best-fitting candidates, needing more fine-grained selection and rejection decisions. Effects related to this process were observed in the LIFG (BA 44/45) and RIFG (BA 45), with greater activation under conditions where higher demands are placed on the evaluation process. This occurs when nonword discrimination involves the rejection of larger sets of active candidates.

The frontal activation associated with cohort competition and selection is broadly consistent with previous findings that bilateral IFG, predominantly on the left, is critically involved in competition and selection processes (Thompson-Schill et al. 1997; Moss et al. 2005; Bozic et al. 2010; Righi et al. 2010; Zhuang et al. 2011). There is also some support for the dissociation that we find between ventral (BA 47/45) and dorsal regions (BA 44/45) for competition and selection processes. Previous studies have shown activation in bilateral ventral IFG (BA 47/45) for competition (e.g. Moss et al. 2005; Bozic et al. 2010; Zhuang et al. 2011), while tasks that emphasize selection processes trigger more activation in BA 44 (e.g. Thompson-Schill et al. 1997). Righi et al. (2010) specifically relate LIFG activity to phonological onset competition—though elicited in a restrictive "visual world" context—and interpret their dorsal BA 44/45 activity as reflecting response-related selection between cohort competitors. This may reflect similar mechanisms to those underpinning cohort selection in the current study, which also implicated a dorsal cluster (BA 44/45).

In addition, Righi et al. (2010) report 2 ventral LIFG clusters (BA 45/47 and insula) that they attribute to semantic/conceptual factors operating in a post-retrieval selection environment, in line with the suggestions of Badre and others (Badre and Wagner 2004; Badre et al. 2005). This analysis does not fit the bilateral BA 45/47 effects seen here (and in Zhuang et al. 2011), which are driven by variations in initial cohort size, and cannot be described as either semantic/conceptual or post-retrieval. These apparent differences in the functional roles assigned to ventral IFG may well reflect the contrasting task demands in these studies and require further research.

Since activity for competition and selection processes only involved bilateral inferior frontal cortices and not other regions, it is unlikely to be driven by controlled retrieval processes (e.g. Wagner et al. 2001), given that both IFG and other regions, such as L temporal and occipital lobes, are commonly coactivated during retrieval processes. Nor are these bilateral IFG activations likely to be related either to working memory load (e.g. Gabrieli et al. 1998) or to maintenance demands of maintaining the activation of word candidates as the sensory input unfolds over time. If these frontal regions were involved in later stages of maintenance, then greater activation would be expected for a low drop-out rate of cohort members, since more word candidates remain in the cohort. This would predict a negative effect of the cohort drop-out rate, which is opposite to the findings here.

The role of posterior (temporal and inferior parietal) regions in lexical access, competition, and selection remains a source of divergence in the literature. In the current research, cohort competition and selection effects are confined to frontal cortex, with temporal regions (L MTG/STG) implicated in lexical retrieval processes associated with the nonword point effect. This L temporal effect reflects dimensions of cohort analysis that do not correlate with the cohort size and drop-out measures used to detect competition and selection effects in frontal regions, but must nevertheless involve processes of discrimination between competing lexical alternatives. It is possible that similar processes were detected in the study by Okada and Hickok (2006), who found effects of lexical neighbourhood density (indirectly related to cohort size) in bilateral posterior STS, with no frontal effects. Prabhakaran et al. (2006) also manipulated neighbourhood density (along with a number of other lexical variables), but found effects more posteriorly in L supramarginal gyrus. Righi et al. (2010) report competitor effects in the same location, along with the frontal effects discussed above. Again, the methodological differences between studies make it difficult to evaluate these contrasting results. In the current study, we saw no trace of competition effects outside bilateral IFG, even at lower thresholds.

In general, this study identifies the neural foundations of an optimally efficient spoken language processing system, functionally differentiated into 3 components, with L anterior and middle MTG/STG (BA 21/22) involved in the online mapping from sounds to meaning, ventral inferior frontal regions (bilateral BA 47 and L BA 45) playing a major role in early, less constrained lexical competition, and dorsal IFG (L BA 44 and bilateral BA 45) more heavily engaged in later fine-grained selection.

References

Allopenna PD, Magnuson JS, Tanenhaus MK. 1998. Tracking the time course of spoken word recognition using eye movements: evidence for continuous mapping models. J Mem Lang. 38:419–439.

Baayen RH, Piepenbroek R, Gulikers L. 1995. The CELEX lexical database. Philadelphia, PA: Philadelphia Linguistic Data Consortium, University of Pennsylvania.

Badre D, Poldrack RA, Pare-Blagoev EJ, Insler RZ, Wagner AD. 2005. Dissociable controlled retrieval and generalized selection mechanisms in ventrolateral prefrontal cortex. Neuron. 47:907–918.

Badre D, Wagner AD. 2004. Selection, integration, and conflict monitoring: assessing the nature and generality of prefrontal cognitive control mechanisms. Neuron. 41:473–487.

Bilenko NY, Grindrod CM, Myers EB, Blumstein SE. 2008. Neural correlates of semantic competition during processing of ambiguous words. J Cogn Neurosci. 21:960–975.

Binder JR, Frost JA, Hammeke TA, Bellgowan PSF, Springer JA, Kaufman JN. 2000. Human temporal lobe activation by speech and nonspeech sounds. Cereb Cortex. 10:512–528.

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

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

Boatman D. 2004. Cortical bases of speech perception: Evidence from functional lesion studies. Cognition. 92:47–65.

Bozic M, Tyler LK, Ives DT, Randall B, Marslen-Wilson WD. 2010. Bihemispheric foundations for human speech comprehension. Proc Natl Acad Sci U S A. 107:17439–17444.

Buchel C, Wise RJS, Mummery CJ, Poline J-B, Friston KJ. 1996. Nonlinear regression in parametric activation studies. Neuroimage. 4:60–66.

Davis MH, Johnsrude IS. 2003. Hierarchical processing in spoken language comprehension. J Neurosci. 23:3423–3431.

Dronkers NF, Wilkins DP, Van Valin RD Jr, Redfern BB, Jaeger JJ. 2004. Lesion analysis of the brain areas involved in language comprehension. Cognition. 92:145–177.

Elman JL, McClelland JL. 1986. Exploiting lawful variability in the speech waveform. In: Perkell JS, Klatt DH, editors. Invariance and variability in speech processes. Hillsdale, NJ: Erlbaum. p. 360–385.

Fahlman SE. 1979. NETL: a system for representing and using real-world knowledge. Cambridge, MA: MIT Press.

Forster KI, Forster JC. 2003. DMDX: a windows display program with millisecond accuracy. Behav Res Methods Instrum Comput. 35:116–124.

Friston KJ, Holmes AP, Worsley KJ, Poline J-P, Frith CD, Frackowiak RSJ. 1995. Statistical parametric maps in functional imaging: a general linear approach. Hum Brain Mapp. 2:189–210.

Gabrieli JDE, Poldrack RA, Desmond JE. 1998. The role of left prefrontal cortex in language and memory. Proc Natl Acad Sci U S A. 95:906–913.

Gaskell MG, Marslen-Wilson WD. 1997. Integrating form and meaning: a distributed model of speech perception. Lang Cogn Process. 12:613–656.

Gaskell MG, Marslen-Wilson WD. 2002. Representation and competition in the perception of spoken words. Cognit Psychol. 45:220–266.

Grindrod CM, Bilenko NY, Myers EB, Blumstein SE. 2008. The role of the left inferior frontal gyrus in implicit semantic competition and selection: an event-related fMRI study. Brain Res. 1229:167–178.

Hall DA, Haggard MP, Akeroyd MA, Palmer AR, Summerfield AQ, Elliott MR, Gurney EM, Bowtell RW. 1999. 'Sparse' temporal sampling in auditory fMRI. Hum Brain Mapp. 7:213–223.

Henson RNA. 2004. Analysis of fMRI timeseries: linear time-invariant models, event-related fMRI and optimal experimental design. In: Frackowiak RSJ, Friston KJ, Frith CD, Dolan RJ, Price CJ, editors. Human brain function. London: Elsevier. p. 793–822.

Hickok G, Poeppel D. 2007. The cortical organization of speech processing. Nature Rev Neurosci. 8:393–402.

Humphries C, Binder JR, Medler DA, Liebenthal E. 2006. Syntactic and semantic modulation of neural activity during auditory sentence comprehension. J Cogn Neurosci. 18:665–679.

Indefrey P, Cutler A. 2004. Pre-lexical and lexical processing in listening. In: Gazzaniga MS, editor. The cognitive neurosciences. 3rd ed. Cambridge (MA): MIT Press. p. 759–774.

January D, Trueswell JC, Thompson-Schill SL. 2008. Co-localization of stroop and syntactic ambiguity resolution in Broca's area: implications for the neural basis of sentence processing. J Cogn Neurosci. 21:2434–2444.

Klatt DH. 1976. Linguistic uses of segmental duration in English: acoustic and perceptual evidence. J Acoust Soc Am. 59:1208–1221.

Lehiste I. 1972. The timing of utterances and linguistic boundaries. J Acoust Soc Am. 51:2018–2024.

Mahon BZ, Costa A, Peterson R, Vargas KA, Caramazza A. 2007. Lexical selection is not by competition: a reinterpretation of semantic interference and facilitation effects in the picture-word interference paradigm. J Exp Psychol Learn Mem Cogn. 33:503–535.

Maldjian JA, Laurienti PJ, Burdette JH. 2004. Precentral gyrus discrepancy in electronic versions of the Talairach atlas. Neuroimage. 21:450–455.

Maldjian JA, Laurienti PJ, Kraft RA, Burdette JH. 2003. An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. Neuroimage. 19:1233–1239.

Marslen-Wilson WD. 1990. Activation, competition and frequency in lexical access. In: Altmann GA, editor. Cognitive models of speech processing: psycholinguistic and computational perspectives. Cambridge (MA): MIT Press. p. 148–172.

Marslen-Wilson WD. 1984. Function and process in spoken word recognition. In: Bouma H, Bouwhuis DG, editors. Attention and performance X: control of language processes. Hillsdale (NJ): Erlbaum. p. 125–150.

Marslen-Wilson WD. 1987. Functional parallelism in spoken word-recognition. Cognition. 25:71–102.

Marslen-Wilson WD. 1975. Sentence perception as an interactive parallel process. Science. 189:226–228.

Marslen-Wilson WD, Tyler LK. 1981. Central processes in speech understanding. Philos Trans R Soc Lond B Biol Sci. 295:317–332.

Marslen-Wilson WD, Welsh A. 1978. Processing interactions and lexical access during word-recognition in continous speech. Cognit Psychol. 10:29–63.

McClelland J, Elman J. 1986. The TRACE model of speech perception. Cognit Psychol. 18:1–86.

Morton J. 1969. Interaction of information in word recognition. Psychol Rev. 76:165–178.

Moss HE, Abdallah S, Fletcher P, Bright P, Pilgrim L, Acres K, Tyler LK. 2005. Selecting among competing alternatives: selection and retrieval in the left inferior frontal gyrus. Cereb Cortex. 15:1723–1735.

Narain C, Scott SK, Wise RJS, Rosen S, Leff A, Iversen SD, Matthews PM. 2003. Defining a left-lateralized response specific to intelligible speech using fMRI. Cereb Cortex. 13:1362–1368.

Norris D. 1994. Shortlist: a connectionist model of continuous speech recognition. Cognition. 52:189–234.

Norris D, McQueen JM. 2008. Shortlist B: a Bayesian model of continuous speech recognition. Psychol Rev. 115:357–395.

Okada K, Hickok G. 2006. Identification of lexical-phonological networks in the superior temporal sulcus using functional magnetic resonance imaging. Neuroreport. 17:1293–1296.

Oldfield RC. 1971. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia. 9:97–113.

Prabhakaran R, Blumstein SE, Myers EB, Hutchison E, Britton B. 2006. An event-related fMRI investigation of phonological-lexical competition. Neuropsychologia. 44:2209–2221.

Price CJ, Wise RJS, Warburton EA, Moore CJ, Howard D, Patterson K, Frackowiak RSJ, Friston KJ. 1996. Hearing and saying: the functional neuro-anatomy of auditory word processing. Brain. 119:919–931.

Ratcliff R. 1993. Methods for dealing with reaction time outliers. Psychol Bull. 114:510–532.

Righi R, Blumstein SE, Mertus J, Worden MS. 2010. Neural systems underlying lexical competition: an eye tracking and fMRI study. J Cogn Neurosci. 22:213–224.

Rodd JM, Davis MH, Johnsrude IS. 2005. The neural mechanisms of speech comprehension: fMRI studies of semantic ambiguity. Cereb Cortex. 15:1261–1269.

Rodd JM, Johnsrude IS, Davis MH. 2012. Dissociating frontotemporal contributions to semantic ambiguity resolution in spoken sentences. Cereb Cortex. 22:1761–1773.

Schnur T, Schwartz MF, Kimberg DY, Hirshorn E, Coslett HB, Thompson-Schill SL. 2009. Localizing interference during naming: convergent neuroimaging and neuropsychological evidence for the function of Broca's area. PNAS. 106:322–327.

Scott SK, Blank CC, Rosen S, Wise RJS. 2000. Identification of a pathway for intelligible speech in the left temporal lobe. Brain. 123:2400–2406.

Scott SK, Wise RJS. 2004. The functional neuroanatomy of prelexical processing in speech perception. Cognition. 92:13–45.

Spitsyna G, Warren JE, Scott SK, Turkheimer FE, Wise RJS. 2006. Converging language streams in the human temporal lobe. J Neurosci. 26:7328–7336.

Thompson-Schill SL, Bedny M, Goldberg RF. 2005. The frontal lobes and the regulation of mental activity. Curr Opin Neurobiol. 15:219–224.

Thompson-Schill SL, D'Esposito M, Aguirre GK, Farah MJ. 1997. Role of left inferior prefrontal cortex in retrieval of semantic knowledge: a reevaluation. Proc Natl Acad Sci U S A. 94:14792–14797.

Tyler LK. 1984. The structure of the initial cohort: evidence from gating. Percept Psychophys. 36:417–427.

Tyler LK, Marslen-Wilson WD. 2008. Fronto-temporal brain systems supporting spoken language comprehension. Philos Trans R Soc Lond B Biol Sci. 363:1037–1054.

Tyler LK, Stamatakis EA, Post B, Randall B, Marslen-Wilson WD. 2005. Temporal and frontal systems in speech comprehension: an fMRI study of past tense processing. Neuropsychologia. 43:1963–1974.

Tyler LK, Voice JK, Moss HE. 2000. The interaction of meaning and sound in spoken word recognition. Psychon Bull Rev. 7:320–326.

Tyler LK, Wessels J. 1983. Quantifying contextual contributions to word-recognition processes. Percept Psychophys. 34:409–420.

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.

Ulrich R, Miller J. 1994. Effects of truncation on reaction-time analysis. J Exp Psychol Gen. 123:34–80.

Wagner AD, Pare-Blagoey EJ, Clark J, Poldrack RA. 2001. Recovering meaning: left prefrontal cortex guides controlled semantic retrieval. Neuron. 31:329–338.

Zhuang J, Randall B, Stamatakis EA, Marslen-Wilson W, Tyler LK. 2011. The interaction of lexical semantics and cohort competition in spoken word recognition: an fMRI study. J Cogn Neurosci. 23:3778–3790.

Zwitserlood P. 1989. The locus of the effects of sentential-semantic context in spoken word processing. Cognition. 32:25–64.