

# Predictive Neural Computations Support Spoken Word Recognition: Evidence from MEG and Competitor Priming

Yingcan Carol Wang,<sup>1</sup> Ediz Sohoglu,<sup>2</sup> Rebecca A. Gilbert,<sup>1</sup>  Richard N. Henson,<sup>1</sup> and  Matthew H. Davis<sup>1</sup>

<sup>1</sup>MRC Cognition and Brain Sciences Unit, University of Cambridge, Cambridge, CB2 7EF, United Kingdom, and <sup>2</sup>School of Psychology, University of Sussex, Brighton, BN1 9RH, United Kingdom

Human listeners achieve quick and effortless speech comprehension through computations of conditional probability using Bayes rule. However, the neural implementation of Bayesian perceptual inference remains unclear. Competitive-selection accounts (e.g., TRACE) propose that word recognition is achieved through direct inhibitory connections between units representing candidate words that share segments (e.g., *hygiene* and *hijack* share /haidʒ/). Manipulations that increase lexical uncertainty should increase neural responses associated with word recognition when words cannot be uniquely identified. In contrast, predictive-selection accounts (e.g., Predictive-Coding) propose that spoken word recognition involves comparing heard and predicted speech sounds and using prediction error to update lexical representations. Increased lexical uncertainty in words, such as *hygiene* and *hijack*, will increase prediction error and hence neural activity only at later time points when different segments are predicted. We collected MEG data from male and female listeners to test these two Bayesian mechanisms and used a competitor priming manipulation to change the prior probability of specific words. Lexical decision responses showed delayed recognition of target words (*hygiene*) following presentation of a neighboring prime word (*hijack*) several minutes earlier. However, this effect was not observed with pseudoword primes (*higent*) or targets (*hijure*). Crucially, MEG responses in the STG showed greater neural responses for word-primed words after the point at which they were uniquely identified (after /haidʒ/ in *hygiene*) but not before while similar changes were again absent for pseudowords. These findings are consistent with accounts of spoken word recognition in which neural computations of prediction error play a central role.

**Key words:** MEG; perception; prediction; priming; speech; STG

## Significance Statement

Effective speech perception is critical to daily life and involves computations that combine speech signals with prior knowledge of spoken words (i.e., Bayesian perceptual inference). This study specifies the neural mechanisms that support spoken word recognition by testing two distinct implementations of Bayes perceptual inference. Most established theories propose direct competition between lexical units such that inhibition of irrelevant candidates leads to selection of critical words. Our results instead support predictive-selection theories (e.g., Predictive-Coding): by comparing heard and predicted speech sounds, neural computations of prediction error can help listeners continuously update lexical probabilities, allowing for more rapid word identification.

## Introduction

In daily conversation, listeners identify ~200 words/minute (Tauroza and Allison, 1990) from a vocabulary of ~40,000 words (Brysbaert et al., 2016). This means that they must recognize 3 or 4 words/s and constantly select from sets of transiently ambiguous words (e.g., *hijack* and *hygiene* both begin with /haidʒ/). Although it is recognized that humans achieve word recognition by combining current speech input with its prior probability using Bayes theorem (Norris and McQueen, 2008; Davis and Scharenborg, 2016), the underlying neural implementation of Bayesian perceptual inference remains unclear (Aitchison and Lengyel, 2017).

Here, we test two computational accounts of spoken word recognition that both implement Bayes rules. In competitive-

Received June 24, 2020; revised May 22, 2021; accepted May 25, 2021.

Author contributions: Y.C.W., R.N.H., and M.H.D. designed research; Y.C.W. performed research; Y.C.W., E.S., and R.A.G. analyzed data; Y.C.W. wrote the first draft of the paper; Y.C.W., E.S., R.A.G., R.N.H., and M.H.D. edited the paper; Y.C.W. and M.H.D. wrote the paper.

This work was supported by UK Medical Research Council SUAG/044 and SUAG/046 G101400, and China Scholarship Council award to Y.C.W. We thank Clare Cook, Ece Kocagoncu, and Tina Emery for assistance with data collection; and Olaf Hauk for advice on MEG data analysis.

The authors declare no competing financial interests.

Correspondence should be addressed to Matthew H. Davis at matt.davis@mrc-cbu.cam.ac.uk or Yingcan Carol Wang at carol.wang@mrc-cbu.cam.ac.uk.

<https://doi.org/10.1523/JNEUROSCI.1685-20.2021>

Copyright © 2021 Wang et al.

This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license, which permits unrestricted use, distribution and reproduction in any medium provided that the original work is properly attributed.

selection accounts (e.g., TRACE; Fig. 1A) (McClelland and Elman, 1986), word recognition is achieved through within-layer lateral inhibition between neural units representing similar words. By this view, *hijack* and *hygiene* compete for identification such that an increase in probability for one word inhibits units representing other similar-sounding words. Conversely, predictive-selection accounts (e.g., Predictive-Coding) (Davis and Sohoglu, 2020) suggest that word recognition is achieved through computations of prediction error (Fig. 1D). On hearing transiently ambiguous speech like /haidʒ/, higher-level units representing matching words make contrasting predictions (/æk/ for *hijack*, /i:n/ for *hygiene*). Prediction error (the difference between sounds predicted and actually heard) provides a signal to update word probabilities such that the correct word can be selected.

In this study, we used the competitor priming effect (Monsell and Hirsh, 1998; Marsolek, 2008), which is directly explicable in Bayesian terms; that is, the recognition of a word (*hygiene*) is delayed if the prior probability of a competitor word (*hijack*) has been increased because of an earlier exposure. This delay could be because of increased lateral inhibition (competitive-selection) or greater prediction error (predictive-selection). Thus, similar behavioral effects of competitor priming are predicted by two distinct neural computations (Spratling, 2008). To distinguish them, it is critical to investigate neural data that reveals the direction, timing, and level of processing at which competitor priming modulates neural responses. Existing neural data remain equivocal with some evidence consistent with competitive-selection (Okada and Hickok, 2006; Bozic et al., 2010), predictive-selection (Gagnepain et al., 2012), or both mechanisms (Brodbeck et al., 2018; Donhauser and Baillet, 2020). We followed these studies in correlating two computational measures with neural activity: lexical entropy (competitive-selection) and segment prediction error (or phoneme surprisal, for predictive-selection).

Here, we used MEG to record the location and timing of neural responses during spoken words recognition in a competitor priming experiment. Pseudowords (e.g., *hijure*) were included in our analysis to serve as a negative control for competitor priming, since existing research found that pseudowords neither produce nor show this effect (Monsell and Hirsh, 1998). We compared items with the same initial segments (words *hygiene*, *hijack*, pseudowords *hijure*, *higent* share /haidʒ/) and measured neural and behavioral effects concurrently to link these two effects for single trials.

While lexical entropy and prediction error are correlated for natural speech, this competitor priming manipulation allows us to make differential predictions as illustrated in Figure 1. Specifically: (1) before the deviation point (DP, the point at which similar-sounding words diverge), competitor priming increases lexical entropy and hence neural responses (Fig. 1B,C; Pre-DP). Such responses might be observed in inferior frontal regions (Zhuang et al., 2011) and posterior temporal regions (Prabhakaran et al., 2006). However, prediction error will be reduced for pre-DP segments, since heard segments are shared and hence more strongly predicted (Fig. 1E,F; Pre-DP). This should be reflected in the superior temporal gyrus (STG) (Sohoglu and Davis, 2016). (2) After the DP, predictive-selection but not competitive-selection accounts propose that pseudowords evoke greater signals in the left-STG, since they evoke maximal prediction errors (Fig. 1E,F; Pseudoword, Post-DP). (3) Furthermore, in predictive-selection theories, competitor priming is associated with an increased STG response to post-DP segments because of enhanced prediction error caused by

mismatch between primed words (predictions) and heard speech (Fig. 1E,F; Word, Post-DP).

## Materials and Methods

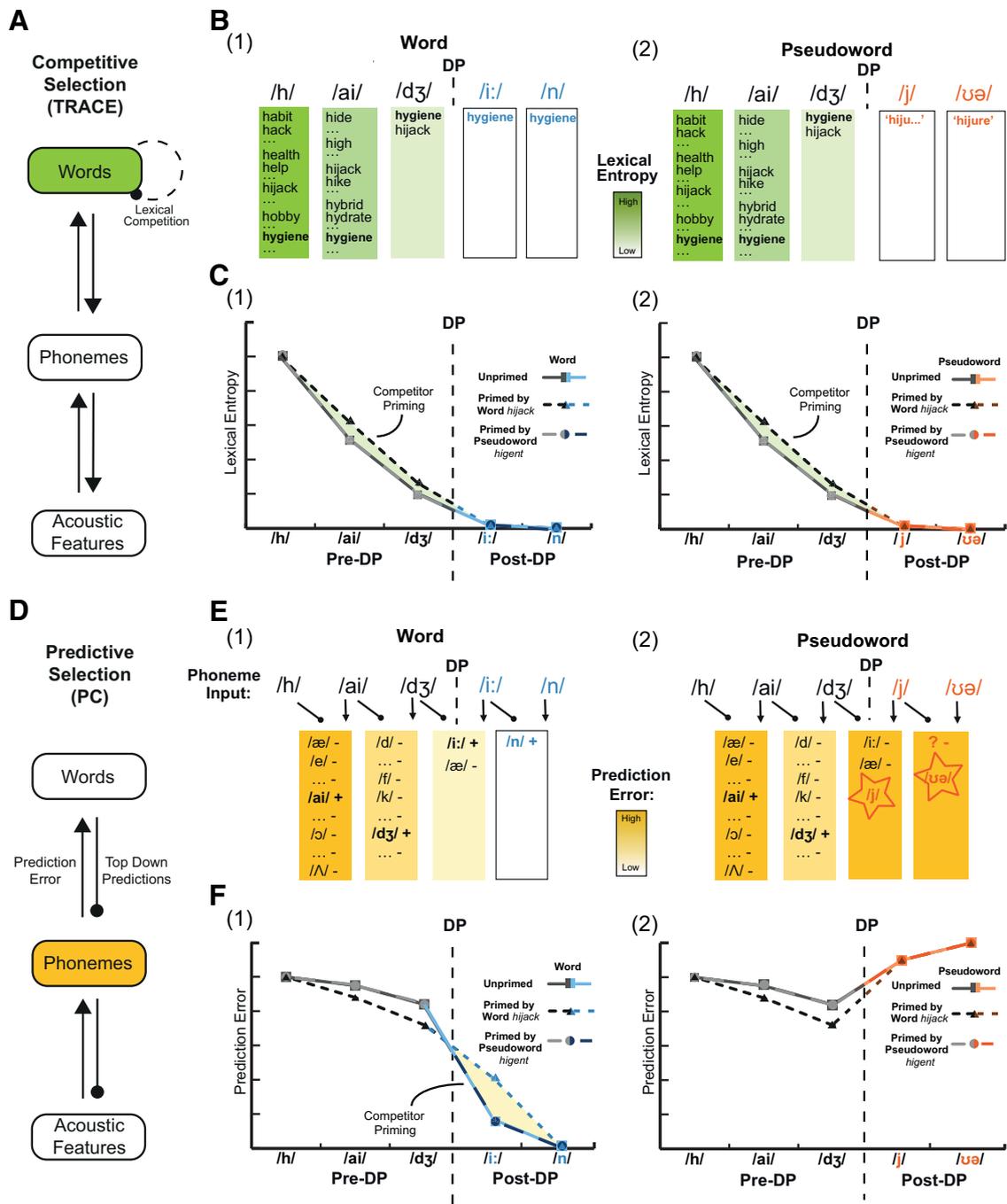
**Participants.** Twenty-four (17 female, 7 male) right-handed, native English speakers were tested after giving informed consent under a process approved by the Cambridge Psychology Research Ethics Committee. This sample size was selected based on previous studies measuring similar neural effects with the same MEG system (Gagnepain et al., 2012; Sohoglu et al., 2012; Sohoglu and Davis, 2016). All participants were 18–40 years of age and had no history of neurologic disorder or hearing impairment based on self-report. Two participants' MEG data were excluded from subsequent analyses, respectively, because of technical problems and excessive head movement, resulting in 22 participants in total. All recruited participants received monetary compensation.

**Experimental design.** To distinguish competitive- and predictive-selection accounts, we manipulated participants' word recognition process by presenting partially mismatched auditory stimuli before targets. Behavioral responses and MEG signals were acquired simultaneously. Prime and target stimuli pairs form a repeated-measures design with two factors (lexicality and prime type). The lexicality factor has 2 levels: word and pseudoword, while the prime type factor contains three levels: unprimed, primed by same lexical status, and primed by different lexical status. Hence the study is a factorial  $2 \times 3$  design with 6 conditions: unprimed word (*hijack*), word-primed word (*hijack-hygiene*), pseudo-word-primed word (*basef-basis*), unprimed pseudoword (*letto*), pseudo-word-primed pseudoword (*letto-lettan*), and word-primed pseudoword (*boycott-boymid*). Prime-target pairs were formed only by stimuli sharing the same initial segments. Items in the two unprimed conditions served as prime items in other conditions and they were compared with target items (Fig. 2A).

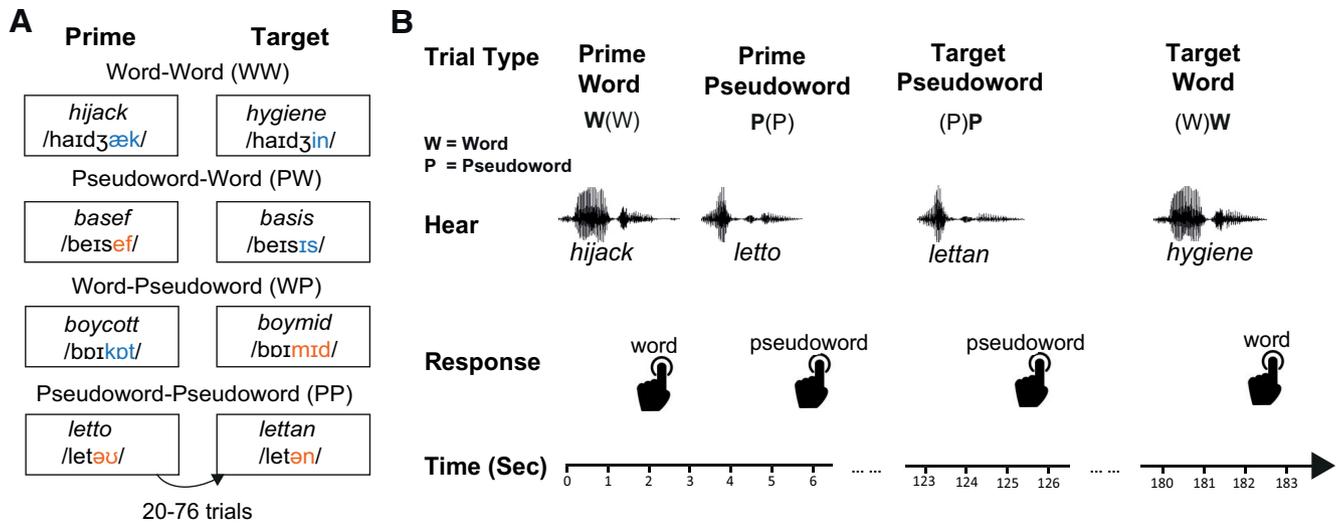
The experiment used a lexical decision task (Fig. 2B) implemented in MATLAB through Psychtoolbox-3 (Kleiner et al., 2007), during which participants heard a series of words and pseudowords while making lexicality judgments to each stimulus by pressing buttons using their left index and middle fingers only, with the index finger pressing one button indicating word and the middle finger pressing the other button indicating pseudoword. A total of 344 trials of unique spoken items were presented every  $\sim 3$  s in two blocks of 172 trials, each block lasting  $\sim 9$  min. Each prime-target pair was separated by 20–76 trials of items that do not start with the same speech sounds, resulting in a relatively long delay of  $\sim 1$ –4 min between presentations of phonologically related items. This delay was chosen based on Monsell and Hirsh (1998), who suggest that it prevents strategic priming effects (Norris et al., 2002). Stimuli from each of the quadruplets were Latin-square counterbalanced across participants; that is, stimulus quadruplets that appeared in one condition for one participant were allocated to another condition for another participant. The stimulus sequences were pseudo-randomized using Mix software (van Casteren and Davis, 2006), so that the same type of lexical status (word/pseudoword) did not appear successively on  $>4$  trials.

**Stimuli.** The stimuli consisted of 160 sets of four English words and pseudowords, with durations ranging from 372 to 991 ms (mean = 643, SD = 106). Each set contained 2 words (e.g., *letter*, *lettuce*) and 2 phonotactically legal pseudowords (e.g., *letto*, *lettan*) that share the same initial segments (e.g., /let/) but diverge immediately afterward.

We used polysyllabic word pairs ( $M_{\text{syllable}} = 2.16$ ,  $SD_{\text{syllable}} = 0.36$ ) instead of monosyllabic ones in our experiments so as to identify a set of optimal lexical competitors that are similar to their prime yet dissimilar from all other items. All words were selected from the CELEX database (Baayen et al., 1993). Their frequencies were taken from SUBTLEX UK corpus (Van Heuven et al., 2014) and restricted to items under 5.5 based on log frequency per million word (Zipf scale) (Van Heuven et al., 2014). In order to ensure that any priming effect was caused purely by phonological but not semantic similarity, we also checked that all prime and target word pairs have a semantic distance of  $>0.7$  on a scale from 0 to 1 based on the Snaut database of semantic similarity scores (Mandera



**Figure 1.** Illustration of neural predictions based on competitive-selection and predictive-selection models, respectively, for recognition of a word (*hygiene*) or pseudoword (*hijure*) that is unprimed or primed by a similar-sounding word (*hijack*) or pseudoword (*higent*). **A**, In a competitive-selection model, such as TRACE (McClelland and Elman, 1986), word recognition is achieved through within-layer lexical competition. **B**, Illustration of the competitive-selection procedure for word (e.g., *hygiene*) and pseudoword (e.g., *hijure*) recognition. Phoneme input triggers the activation of multiple words beginning with the same segments, which compete with each other until one word is selected. No word can be selected when hearing a pseudoword, although it would be expected that lexical probability (although not lexical entropy) should be greater for words than for pseudowords. **C**, Illustration of neural predictions based on lexical entropy. Lexical entropy gradually reduces to zero as more speech is heard. Before the deviation point (DP) at which the prime (*hijack*) and target (*hygiene*) diverge, these items are indistinguishable, and competitor priming should transiently increase lexical entropy (shaded area). After the DP, competitor priming should not affect entropy since prime and target words can be distinguished. **D**, In a predictive-selection model, such as the Predictive-Coding account (PC) (Davis and Sohoglu, 2020), words are recognized by minimizing prediction error, which is calculated by subtracting the predicted segments from the current sensory input. **E**, Illustration of the predictive-selection procedure during word (e.g., *hygiene*) and pseudoword (e.g., *hijure*) recognition. Speech input evokes predictions for the next segment (based on word knowledge as in **B**), which is then subtracted from the speech input and used to generate prediction errors that update lexical predictions: +, confirmed predictions that increase lexical probability; -, disconfirmed predictions that decrease lexical probability. **F**, Illustration of neural predictions based on segment prediction error. Before the DP, priming of initial word segments should strengthen predictions and reduce prediction error. There will also be greater mismatch between predictions and heard speech for competitor-primed words; hence, primed words should evoke greater prediction error than unprimed words (shaded area). This increased prediction error should still be less than that observed for pseudowords, which should evoke maximal prediction error regardless of competitor priming because of their post-DP segments being entirely unpredictable.



**Figure 2.** Experimental design. **A**, Four different types of prime-target pairs. Each pair was formed by two stimuli from the same quadruplet, separated by between 20 and 76 trials of items that do not share the same initial segments. **B**, Lexical decision task. Participants made lexicality judgments to each item they heard via left-hand button-press. The RT was recorded from the onset of the stimuli. As shown, items within each quadruplet are repeated after a delay of ~1-4 min following a number of other intervening stimuli.

et al., 2017), such that morphologic relatives (e.g., darkly/darkness) were excluded.

All spoken stimuli were recorded onto a Marantz PMD670 digital recorder by a male native speaker of southern British English in a sound-isolated booth at a sampling rate of 44.1 kHz. Special care was taken to ensure that shared segments of stimuli were pronounced identically (any residual acoustic differences were subsequently eliminated using audio morphing as described below).

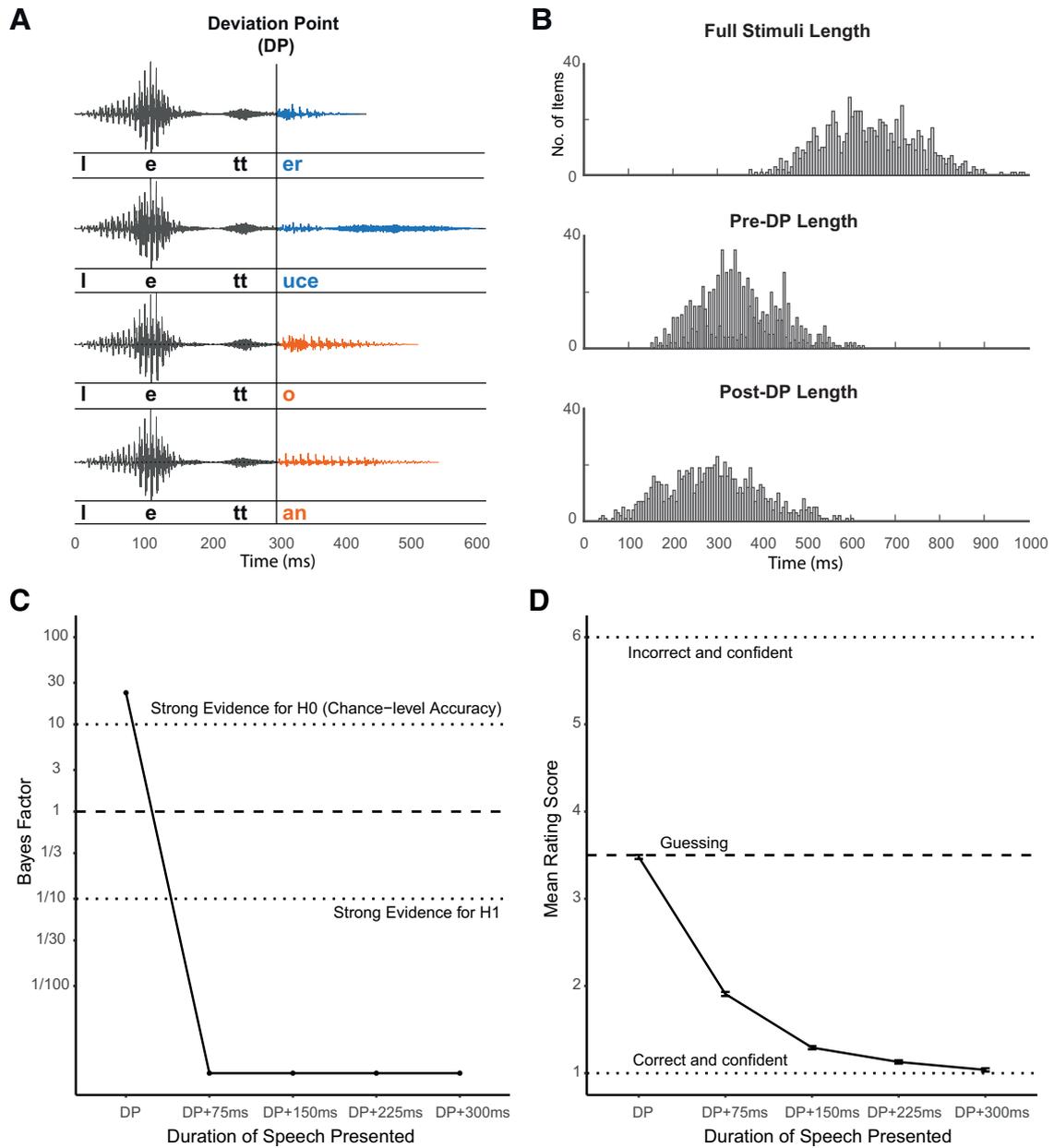
The point when items within each quadruplet begin to acoustically differ from each other is the DP (Fig. 3A). Pre-DP length ranged from 150 to 672 ms (mean = 353, SD = 96), while post-DP length ranged from 42 to 626 ms (mean = 290, SD = 111; Fig. 3B). Epochs of MEG data were time-locked to the DP. Using phonetic transcriptions (phonDISC) in CELEX, the location of the DP was decided based on the phoneme segment at which items within each quadruplet set diverge ( $M_{\text{seg}} = 3.53$ ,  $SD_{\text{seg}} = 0.92$ ). To determine when in the speech files correspond to the onset of the first post-DP segment, we aligned phonetic transcriptions to corresponding speech files using the WebMAUS forced alignment service (Schiel, 1999; Kisler et al., 2017). In order to ensure that the pre-DP portion of the waveform was acoustically identical, we cross-spliced the pre-DP segments of the four stimuli within each quadruplet and conducted audio morphing to combine the syllables using STRAIGHT (Kawahara, 2006) implemented in MATLAB. This method decomposes speech signals into source information and spectral information, and permits high quality speech resynthesis based on modified versions of these representations. This enables flexible averaging and interpolation of parameter values that can generate acoustically intermediate speech tokens (see, e.g., Rogers and Davis, 2017). In the present study, this method enabled us to present speech tokens with entirely ambiguous pre-DP segments, and combine these with post-DP segments without introducing audible discontinuities or other degradation in the speech tokens. This way, phonological co-articulation in natural speech was reduced to the lowest level possible at the DP; hence, any cross-stimulus divergence evoked in neural responses can only be caused by post-DP deviation.

**Post-test gating study.** As encouraged by a reviewer, we conducted a post-test perceptual experiment using a gating task to confirm that the cross-splicing and morphing of our stimuli worked as expected. This experiment used a gating task implemented in JavaScript through JPsych (de Leeuw, 2015). During the experiment, auditory segments of all 160 pairs of words used in the MEG study were played. Twenty British English speakers were recruited through Prolific Academic online with monetary compensation. The sample size was selected based on a similar gating study conducted by Davis et al. (2002). Participants were

evenly divided into two groups, one group were presented with 160 stimuli words with different pre-DP segments (e.g., *hygiene*), while the other group were presented with the other paired 160 stimuli (e.g., *hijack*). Therefore, participants only ever heard one of the two items in each pair. Stimuli segments of each word item consist of the pre-DP segment and, depending on the stimuli length, also longer segments that are 75, 150, 225, and 300 ms after DP. The segments of each word were presented in a gating manner, with the shortest segment played the first and the full item played at the end. After hearing each segment (e.g., /haɪdʒ/), participants were also presented with the writing of the word (e.g., *hygiene*) that contained the segment and the other paired word that shared the same pre-DP segment (e.g., *hijack*) on the screen. We asked the participants to choose which item the auditory segment matches and indicate their confidence from a rating scale of 1 to 6, with 1 representing being very confident that the item is the one on the left and 6 representing being very confident that the item is the one on the right, while 3 and 4 represent guessing the possible item. In order to avoid potential practice effect, we also added 40 filler stimuli that are identifiable on initial presentation.

Given our goal of assessing whether there is any information to distinguish the words before the divergence point, we needed to adopt an analysis approach that could confirm the null hypothesis that no difference exists between perception of the shared first syllable of word pairs like *hijack* and *hygiene*. We therefore analyzed the results using Bayesian methods which permit this inference. Participants' response accuracy was analyzed using mixed-effect logistic regression, and confidence rating scores were analyzed using mixed-effect linear regression using the brms package (Bürkner, 2017) implemented in R. Response scores were computed in a way such that correct and most confident responses were scored 1, while incorrect and most confident responses were scored 6 and so on. Participants and items were included as random factors of the models, and there was no fixed factor since we are only interested in the intercepts, whose estimates indicate the logit transformed proportion of correctness in the logistic model and the mean rating in the linear model, respectively. We chose weakly informative priors for each model and conducted Bayes factor analyses through the Savage-Dickey density ratio method (Wagenmakers et al., 2010). Model estimate, SE, lower and upper boundary of 95% credible interval (CI) are also reported.

When checking our data, we found that 16 pairs of word items were not morphed correctly; hence, the spectral information of the pre-DP segments of these word pairs was not exactly the same, and some of them diverged acoustically before the DP because of coarticulation. Therefore, we excluded these items from analyses of the gating data and confirmed that excluding these items did not modify the interpretation or significance of the MEG or behavioral results reported in the paper.



**Figure 3.** Stimuli and post-test gating study results. **A**, Stimuli within the same quadruplet have identical onsets in STRAIGHT parameter space (Kawahara, 2006) and thus only diverge from each other after the DP. MEG responses were time-locked to the DP. **B**, Stimuli length histogram. **C**, Bayes factor for chance level accuracy (BF01) at each post-DP alignment point of the stimuli in the post-test gating study. **D**, Mean rating score at each post-DP alignment point of the stimuli in the gating study.

As shown in Figure 3C, we found that when gating segments ended at the DP, Bayes factor provides strong evidence in favor of the null hypothesis, chance-level accuracy (i.e., proportion of correct responses is 0.5),  $\beta = 0.04$ ,  $SE = 0.08$ ,  $ICI = -0.11$ ,  $uCI = 0.20$ ,  $BF01 = 23.04$ . This indicates that participants could not predict the full stimuli based on hearing the pre-DP segments. On the other hand, the Bayes factor at later alignment points is close to 0, providing extremely strong evidence for the alternative hypothesis that the proportion of correct responses is higher than 0.5 (75 ms post-DP:  $\beta = 3.41$ ,  $SE = 0.22$ ,  $ICI = 2.99$ ,  $uCI = 3.85$ ,  $BF01 < 0.01$ ; 150 ms post-DP:  $\beta = 6.26$ ,  $SE = 0.56$ ,  $ICI = 5.24$ ,  $uCI = 7.41$ ,  $BF01 < 0.01$ ; 225 ms post-DP:  $\beta = 7.39$ ,  $SE = 1.02$ ,  $ICI = 5.65$ ,  $uCI = 9.72$ ,  $BF01 < 0.01$ ; 300 ms post-DP:  $\beta = 8.04$ ,  $SE = 1.88$ ,  $ICI = 4.99$ ,  $uCI = 12.32$ ,  $BF01 < 0.01$ ). Figure 3D shows that, with the gating segment becoming longer, the rating scores gradually reduce (lower scores indicating more accurate and more confident identification). We examined whether the mean score at the DP is equal to 3.5 (i.e., chance performance) and found strong evidence supporting the null hypothesis,  $\beta = -0.02$ ,  $SE = 0.04$ ,  $ICI = -0.10$ ,  $uCI = 0.06$ ,  $BF01 = 21.79$ , which is

consistent with the accuracy results. Furthermore, to refine the estimate of the time point at which participants recognize the stimuli with enough confidence, we also investigated at what alignment point is there evidence showing the mean score  $< 2$  (i.e., participants indicating more confident identification). We found moderate evidence supporting the null hypothesis (mean score equals to 2) at 75 ms post-DP ( $\beta = -0.09$ ,  $SE = 0.08$ ,  $ICI = -0.25$ ,  $uCI = 0.07$ ,  $BF01 = 6.07$ ), but extremely strong evidence in favor of the alternative hypothesis at 150 ms post-DP ( $\beta = -0.71$ ,  $SE = 0.05$ ,  $ICI = -0.79$ ,  $uCI = 0.62$ ,  $BF01 < 0.01$ ). These results show that critical acoustic information that supports confident word recognition arrives between 75 and 150 ms post-DP.

Overall, the post-test gating study confirmed that the pre-DP segments of correctly morphed stimuli are not distinguishable within each stimuli set. However, since we found items that were not correctly morphed during this control study, we did a thorough check of our stimuli and identified all the items with pre-DP acoustic differences (16 words and 12 pseudowords), which resulted in 8.68% of all trials presented in the MEG study. In order to double check our MEG study

results, we then removed all these problematic trials from the data and reanalyzed the data using the same methods as described in Materials and Methods. Fortunately, we did not find any inconsistent pattern or significance in our behavioral or neural results compared with those reported with all trials included (see Tables 1–5). Therefore, we kept the original MEG and behavioral results with all items included in this paper.

**Behavioral data analyses.** Response times (RTs) were measured from the onset of the stimuli and inverse-transformed so as to maximize the normality of the data and residuals; figures report untransformed RTs for clarity. Inverse-transformed RTs and error rates were analyzed using linear and logistic mixed-effect models, respectively, using the lme4 package in R (Bates et al., 2014). Lexicality (word, pseudoword) and prime type (unprimed, primed by same lexical status, primed by different lexical status) were fixed factors, while participant and item were random factors. Maximal models accounting for all random effects were attempted wherever possible, but reduced random effects structures were applied when the full model did not converge (Barr et al., 2013). Likelihood-ratio tests comparing the full model to a nested reduced model using the  $\chi^2$  distribution were conducted to evaluate main effects and interactions. Significance of individual model coefficients were obtained using  $t$  (reported by linear mixed-effect models) or  $z$  (reported by logistic mixed-effect models) statistics in the model summary. One-tailed  $t$  statistics for RTs are also reported for two planned contrasts: (1) word-primed versus unprimed conditions for word targets, and (2) word-primed versus pseudoword-primed conditions for word targets.

When assessing priming effects, we excluded data from target trials in which the participant made an error in the corresponding prime trial, because it is unclear whether such target items will be affected by priming given that the prime word was not correctly identified. In addition, three trials with RTs shorter than the average pre-DP length (353 ms) were removed from further analysis, since responses before words and pseudowords acoustically diverge are too quick to be valid lexical decision responses.

**MEG data acquisition, processing, and analyses.** Magnetic fields were recorded with a VectorView system (Elekta Neuromag) which contains a magnetometer and two orthogonal planar gradiometers at each of 102 locations within a hemispherical array around the head. Although electric potentials were recorded simultaneously using 68 Ag-AgCl electrodes according to the extended 10–10 system, these EEG data were excluded from further analysis because of excessive noise. All data were digitally sampled at 1 kHz. Head position was monitored continuously using five head-position indicator coils attached to the scalp. Vertical and horizontal electro-oculograms were also recorded by bipolar electrodes. A 3D digitizer (FASTRAK; Polhemus) was used to record the positions of three anatomic fiducial points (the nasion, left and right preauricular points), head-position indicator coils, and evenly distributed head points for use in source reconstruction.

MEG data were preprocessed using the temporal extension of Signal Source Separation in MaxFilter software (Elekta Neuromag) to reduce noise sources, normalize the head position over blocks and participants to the sensor array, and reconstruct data from bad MEG sensors. Subsequent processing was conducted in SPM12 (<https://www.fil.ion.ucl.ac.uk/spm/>) and FieldTrip (<http://www.fieldtriptoolbox.org/>) software implemented in MATLAB. The data were epoched from –1100 to 2000 ms time-locked to the DP and baseline corrected relative to the –1100 to –700 ms before the DP, which is a period before the onset of speech for all stimuli (Fig. 1C). Low-pass filtering to 40 Hz was conducted both before and after robust averaging across trials (Litvak et al., 2011). A time window of –150 to 0 ms was defined for pre-DP comparisons based on the shortest pre-DP stimuli length. A broad window of 0–1000 ms was defined for post-DP comparisons, which covered the possible period for lexicality and prime effects. After averaging over trials, an extra step was taken to combine the gradiometer data from each planar sensor pair by taking the root-mean square of the two amplitudes.

Sensor data from magnetometers and gradiometers were analyzed separately. We converted the sensor data into 3D images (2D sensor  $\times$  time) and performed  $F$  tests for main effects across sensors and time (the term “sensors” denotes interpolated sensor locations in 2D image space).

Reported effects were obtained with a cluster-defining threshold of  $p < 0.001$ , and significant clusters identified as those whose extent (across space and time) survived  $p < 0.05$  familywise error (FWE) correction using Random Field Theory (Kilner and Friston, 2010). ROI analyses for the priming effect were then conducted over sensors and time windows that encompassed the significant pseudoword  $>$  word cluster, orthogonal to priming effects. When plotting waveforms and topographies, data are shown for sensors nearest to the critical points in 2D image space.

Apart from the two planned contrasts mentioned above (see Behavioral data analyses), which were applied to post-DP analysis, one-tailed  $t$  statistics was also reported on the pre-DP planned contrast between unprimed and word-primed items.

**Source reconstruction.** In order to determine the underlying brain sources underlying the sensor-space effects, source reconstruction was conducted using SPM’s Parametric Empirical Bayes framework (Henson et al., 2011). To begin with, we obtained T1-weighted structural MRI scans from each participant on a 3T Prisma system (Siemens) using an MPRAGE sequence. The scan images were segmented and normalized to an MNI template brain in MNI space. The inverse of this spatial transformation was then used to warp canonical meshes derived from that template brain back to each subject’s MRI space (Mattout et al., 2007). Through this procedure, canonical cortical meshes containing 8196 vertices were generated for the scalp and skull surfaces. We coregistered the MEG sensor data into the structural MRI space for each participant by using their respective fiducials, sensor positions, and head-shape points (with nose points removed because of the absence of the nose on the T1-weighted MRI). Using the single-shell model, the lead field matrix for each sensor was computed for a dipole at each canonical cortical mesh vertex, oriented normal to the local curvature of the mesh.

Source inversion was performed with all conditions pooled together using the IID solution, equivalent to classical minimum norm, fusing the magnetometer and gradiometer data (Henson et al., 2011). The resulting inversion was then projected onto wavelets spanning frequencies from 1 to 40 Hz and from –150 to 0 ms time samples for pre-DP analysis and 400–900 ms for post-DP analysis. This post-DP time window was defined by overlapping temporal extent of the pseudoword  $>$  word cluster between gradiometers and magnetometers. The total energy within these time-frequency windows was summarized by taking the sum of squared amplitudes, which was then written to 3D images in MNI space.

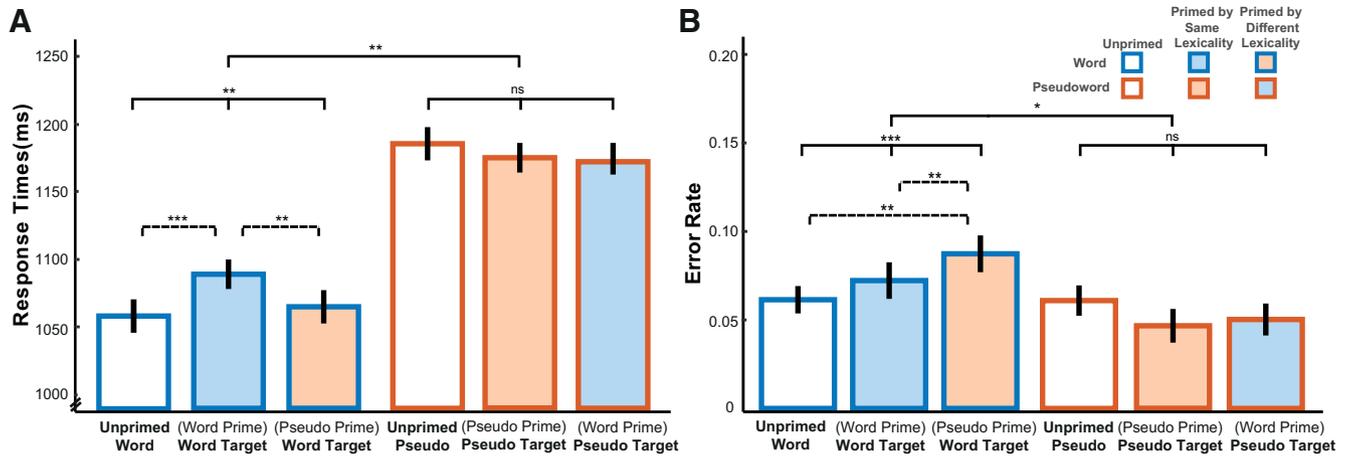
Reported effects for source analyses were obtained with a cluster-defining threshold of  $p < 0.05$  (FWE-corrected). And as in sensor space, ROI analyses were conducted over significant sensors and time windows from the orthogonal pseudoword  $>$  word cluster. Factorial ANOVA were conducted on main effects and one-tailed paired  $t$  tests on planned contrasts (see MEG data acquisition and processing).

## Results

### Behavior

#### RTs

As shown in Figure 4A and Table 1, factorial analysis of lexicality (word, pseudoword) and prime type (unprimed, primed by same lexical status, primed by different lexical status) indicated a significant main effect of lexicality, in which RTs for pseudowords were significantly longer than for words,  $\chi^2_{(3)} = 23.60$ ,  $p < 0.001$ . In addition, there was a significant interaction between lexicality and prime type,  $\chi^2_{(2)} = 10.73$ ,  $p = 0.005$ . This interaction was followed up by two separate one-way models for words and pseudowords, which showed a significant effect of prime type for words,  $\chi^2_{(2)} = 10.65$ ,  $p = 0.005$ , but not for pseudowords,  $\chi^2_{(2)} = 1.62$ ,  $p = 0.445$ . Consistent with the competitor priming results from Monsell and Hirsh (1998), words that were primed by another word sharing the same initial segments were recognized significantly more slowly than unprimed words (for mean raw RTs, see Fig. 3A),  $\beta = 0.02$ ,  $SE = 0.01$ ,  $t_{(79.69)} = 3.33$ ,  $p < 0.001$ , and more slowly than pseudoword-primed words,



**Figure 4.** RT results (A) and accuracy results (B) of the lexical decision task. Bars are color-coded by lexicality and prime type on the x axis (words, blue frame; pseudowords, orange frame; unprimed, no fill; primed by same lexicality, consistent fill and frame colors; primed by different lexicality, inconsistent fill and frame colors). Bars represent the subject grand averages. Error bars indicate  $\pm$  within-subject SE, adjusted to remove between-subjects variance (Cousineau, 2005). Statistical significance is shown based on generalized linear mixed-effects regression: \* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\* $p < 0.001$ . Statistical comparisons shown with solid lines indicate the lexicality by prime-type interaction and main effects of prime-type for each lexicality, whereas comparisons with broken lines indicate the significance of pairwise comparisons.

**Table 1.** Behavioral RT analyses on all data versus data excluding items with pre-DP acoustic differences (identified in the gating post-test)<sup>a</sup>

Contrast	All data			Data with exclusion		
	$\chi^2$	$t$	$p$	$\chi^2$	$t$	$p$
Lexicality	23.60	—	<0.001	28.87	—	<0.001
Lexicality $\times$ prime type	10.73	—	0.005	8.52	—	0.014
Word prime type	10.65	—	0.005	8.57	—	0.014
Word-word > word	—	3.33	<0.001	—	3.00	0.002
Word-word > pseudo-word	—	2.37	0.009	—	2.30	0.011
Pseudo prime type	1.62	—	0.445	0.65	—	0.720

<sup>a</sup>Reported pairwise effects (planned) are one-tailed. Word-word, word-primed word; pseudo-word, pseudo-word-primed word.

**Table 2.** Behavioral accuracy analyses on all data versus data excluding items with pre-DP acoustic differences<sup>a</sup>

Contrast	All data			Data with exclusion		
	$\chi^2$	$t$	$p$	$\chi^2$	$t$	$p$
Lexicality	7.31	—	0.063	6.40	—	0.094
Lexicality $\times$ prime type	6.08	—	0.048	6.98	—	0.031
Word prime type	13.95	—	<0.001	14.97	—	<0.001
Pseudo-word > word	—	3.14	0.005	—	3.03	0.007
Pseudo-word > word-word	—	3.07	0.007	—	3.05	0.007
Pseudo prime type	1.93	—	0.381	3.16	—	0.206

<sup>a</sup>Reported pairwise effects are Bonferroni-corrected. Word-word, word-primed word; pseudo-word, pseudo-word-primed word.

$\beta = 0.02$ ,  $SE = 0.01$ ,  $t_{(729,89)} = 2.37$ ,  $p = 0.009$ . As mentioned earlier (see Introduction), both competitive- and predictive-selection models predicted longer RTs to word-primed target words compared with unprimed words, it is hence critical to distinguish the two accounts through further investigation of the MEG responses.

**Accuracy**

Figure 4B and Table 2 indicate that there was a trend toward more lexical decision errors in response to words than to pseudowords, although this lexicality effect was marginal,  $\chi^2_{(3)} = 7.31$ ,  $p = 0.063$ . The error rates for words and pseudowords were also affected differently by priming, as indicated by a significant

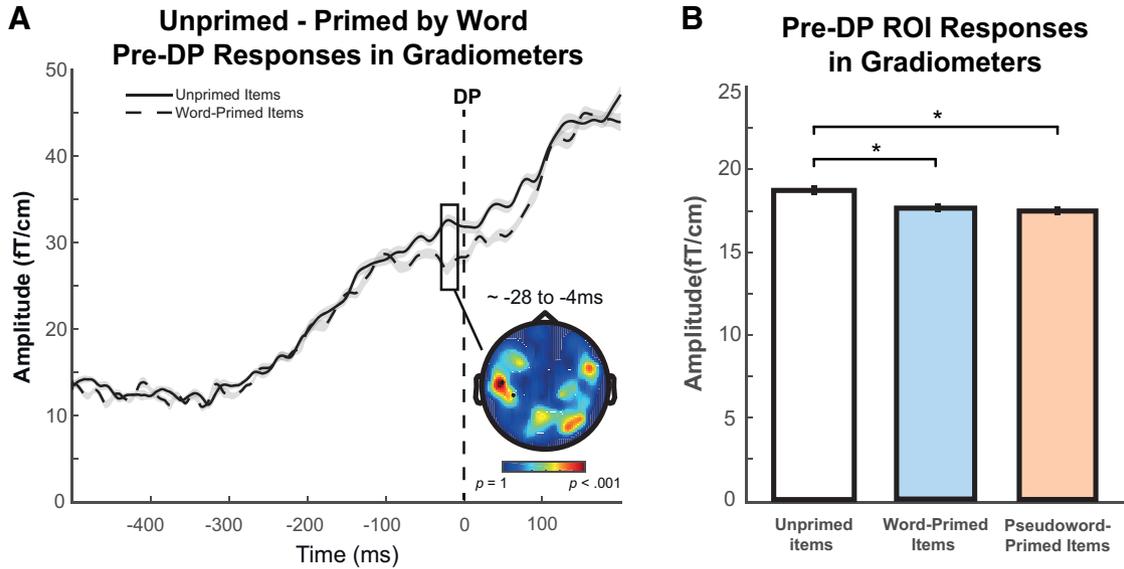
interaction between lexicality and prime type,  $\chi^2_{(2)} = 6.08$ ,  $p = 0.048$ . Follow-up analyses using two separate models for each lexicality type showed there was a main effect of prime type for words,  $\chi^2_{(2)} = 13.95$ ,  $p < 0.001$ , but not for pseudowords,  $\chi^2_{(2)} = 1.93$ ,  $p = 0.381$ . Since we had not anticipated these priming effects on accuracy, *post hoc* pairwise  $z$  tests were Bonferroni-corrected for multiple comparisons. These showed that pseudo-word priming reliably increased the error rate compared with the unprimed condition,  $\beta = 1.68$ ,  $SE = 0.54$ ,  $z = 3.14$ ,  $p = 0.005$ , and to the word-primed condition,  $\beta = 2.74$ ,  $SE = 0.89$ ,  $z = 3.07$ ,  $p = 0.007$ . Although no specific predictions on accuracy were made *a priori* by either competitive- or predictive-selection model, it is worth noting that participants might have expected pseudowords to be repeated given the increased error rate of responses to pseudo-word-primed target words.

**MEG**

In order to explore the impact of lexicality and competitor priming on neural responses to critical portions of speech stimuli, both before and after they diverge from each other, MEG responses were time-locked to the DP. All reported effects are FWE-corrected at cluster level for multiple comparisons across scalp locations and time at a threshold of  $p < 0.05$ . We reported data from gradiometers, magnetometers, and source space wherever possible, since sensor  $\times$  time analyses help define the time windows used by source localization. Although some minor effects were shown in only one of these analyses, our most interesting effects are reliable in all three data types.

**Pre-DP analyses**

We assessed neural responses before the DP, during which only the shared speech segments have been heard and hence the words and pseudowords in each stimulus set are indistinguishable. Since there could not have been any effect of lexical status pre-DP, only prime type effects were considered in this analysis. Predictive- and competitive-selection accounts make opposite predictions for pre-DP neural signals evoked by word-primed items compared with unprimed items. We therefore conducted an  $F$  test for neural differences between these two conditions across the scalp and source spaces over a time period of  $-150$  to  $0$  ms before the DP. A significant cluster of 295 sensor  $\times$  time



**Figure 5.** Pre-DP results. **A**, Pre-DP response difference between items that are unprimed and primed by a word in MEG gradiometer sensors within  $-150$  to  $0$  ms (a time window at which words and pseudowords are indistinguishable). The topographic plot represents  $F$  statistics for the entire sensor array with scalp locations that form a statistically significant cluster highlighted and marked with black dots. Waveforms represent MEG response averaged over the spatial extent of the significant cluster shown in the topography. Gray shade of waveforms represents  $\pm$  within-participant SE, adjusted to remove between-participants variance (Cousineau, 2005). **B**, ROI analysis of neural responses evoked by unprimed and primed items averaged over the same pre-DP time period of  $-150$  to  $0$  ms but across gradiometer sensor locations which showed the post-DP pseudoword > word lexicality effect (see Fig. 6A). Bars are color-coded by prime type on the x axis (unprimed items, no fill; word-primed items, blue; pseudoword-primed items, orange; black frame indicates that words and pseudowords are indistinguishable). Error bars indicate  $\pm$  within-participant SE, adjusted to remove between-participant variance. \* $p < 0.05$ .

**Table 3. Pre-DP MEG analyses of unprimed > word-primed items and post-DP MEG analyses of pseudoword > word on all data versus data excluding items with pre-DP acoustic differences<sup>a</sup>**

Time window	Modality	All data			Data with exclusion		
		$P_{FWE-corr}$	Cluster size	Latency (ms)	$P_{FWE-corr}$	Cluster size	Latency (ms)
Pre-DP	Grad	0.023	295	$-28$ to $-4$	0.005	426	$-25$ - $0$
Post-DP	Grad	<0.001	39335	313-956	<0.001	30811	320-775
	Mag	<0.001	68517	359-990	<0.001	69777	362-988
	Source	<0.001	2315	400-900	<0.001	2287	400-900

<sup>a</sup>Reported effects are FWE-corrected at cluster level at  $p < 0.05$ . Grad, gradiometers; Mag, magnetometers.

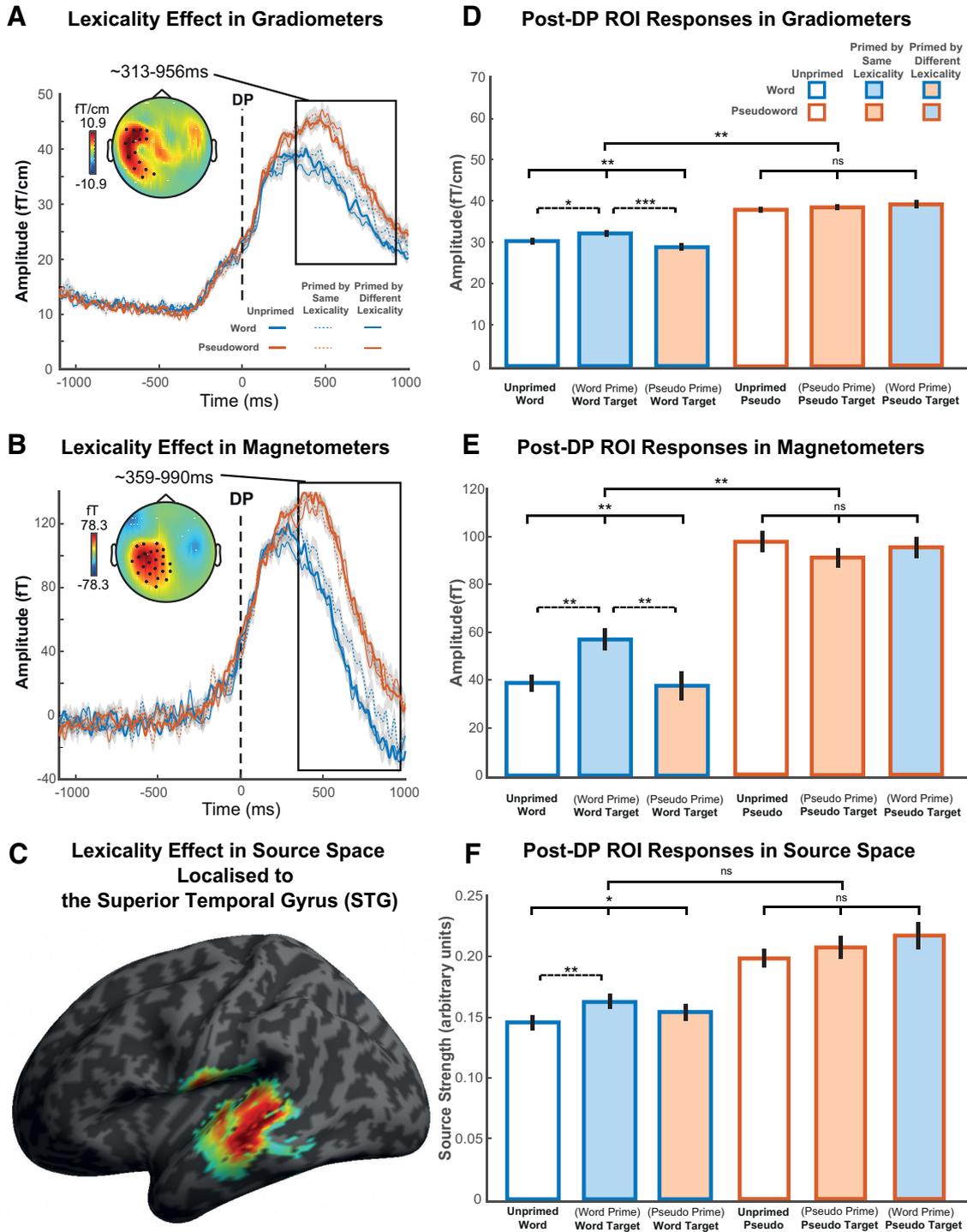
points ( $p = 0.023$ ) was found in gradiometers over the mid-left scalp locations from  $-28$  to  $-4$  ms (Fig. 5A; Table 3), in which unprimed items evoked significantly greater neural responses than word-primed items. On the suggestion of a reviewer, and mindful of the potential for these pre-DP neural responses to be modulated by post-DP information, we report an additional analysis with a lengthened analysis time window of  $-150$  to  $100$  ms. Again, we found a significant unprimed > word-primed cluster of 313 sensor  $\times$  time points ( $p = 0.033$ ) over the exact same locations in gradiometers from  $-28$  to  $-3$  ms pre-DP, which confirmed that this pre-DP effect was not pushed forward by any post-DP effect. We did not find any cluster showing stronger neural responses for word-primed items than unprimed items, and no clusters survived correction for multiple comparisons for magnetometer responses or for analysis in source space.

To further examine these results, we also conducted ROI analysis of gradiometer signals evoked by unprimed and primed items averaged over the same  $-150$  to  $0$  ms pre-DP time window but across the scalp locations that showed the post-DP lexicality effect at which pseudowords elicited greater neural responses

than words (Fig. 6A). As shown in Figure 5B and Table 4, the results indicated that unprimed items elicited significantly stronger neural responses than word-primed items,  $t_{(21)} = 2.41$ ,  $p = 0.013$ , consistent with the whole-brain analysis. In particular, the cluster shown in Figure 5A partially overlaps with the post-DP pseudoword > word cluster in Figure 6A. The direction and location of these pre-DP neural responses are in accordance with the predictive-selection account and inconsistent with the competitive-selection account. A surprising finding is that *post hoc* analysis also showed greater neural responses evoked by unprimed items than pseudoword-primed items,  $t_{(21)} = 2.69$ ,  $p = 0.014$ , although we had not predicted these effects from pseudoword primes.

*Post-DP analyses*

We then examined the post-DP response differences between words and pseudowords (lexicality effect). The gradiometer sensors showed a significant cluster of 39,335 sensor  $\times$  time points ( $p < 0.001$ ) over the left side of the scalp at 313-956 ms post-DP (Fig. 6A; Table 3). In this cluster, pseudowords evoked a significantly stronger neural response than words. Similarly, magnetometer sensors also detected a significant left-hemisphere cluster of 68,517 sensor  $\times$  time points ( $p < 0.001$ ) at 359-990 ms post-DP (Fig. 6B; Table 3) showing the same lexicality effect. We did not find any significant cluster in which words evoked greater neural responses than pseudowords. These results are consistent with findings from Gagnepain et al. (2012). To locate the likely neural source of the effects found in sensor space, we conducted source reconstruction by integrating gradiometers and magnetometers. As shown in Figure 6C, results from source space showed that neural generators of the lexicality effect were estimated to lie within the left STG (volume of 2315 voxels,  $p < 0.001$ , peak at  $x = -46$ ,  $y = -36$ ,  $z = 0$ ;  $x = -52$ ,  $y = -34$ ,  $z =$



**Figure 6.** Post-DP results showing lexically effects and corresponding ROI responses evoked by conditions of interest. **A, B**, Post-DP lexicality effects in MEG gradiometer and magnetometer sensors. The topographic plots represent the statistically significant cluster with a main effect of lexicality (pseudoword > word). Waveforms represent MEG response averaged over the spatial extent of the significant cluster shown in the topography. Gray shade of waveforms represents  $\pm$  within-participant SE, adjusted to remove between-participants variance. **C**, Statistical parametric map showing the cluster (pseudoword > word) rendered onto an inflated cortical surface of the MNI standard brain thresholded at FWE-corrected cluster-level  $p < 0.05$ , localized to the left STG. **D–F**, Post-DP ROI ANOVA on neural signals and source strength evoked by conditions of interest averaged over the time window and scalp locations of the significant cluster shown in **A–C**. Bars are color-coded by lexicality and prime type on the x axis (words, blue frame; pseudowords, orange frame; unprimed, no fill; primed by same lexicality, consistent fill and frame colors; primed by different lexicality, inconsistent fill and frame colors). Error bars indicate  $\pm$  within-participant SE, adjusted to remove between-participants variance. Statistical significance from ANOVAs: \* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\* $p < 0.001$ . Statistical comparisons shown with solid lines indicate the lexicality by prime-type interaction and main effects of prime-type for each lexicality, whereas comparisons with broken lines indicate the significance of planned pairwise comparisons.

–6;  $x = -56$ ,  $y = -28$ ,  $z = -10$ ). This location, and direction of response, is consistent with a sublexical (e.g., phonemic) process being modulated by lexicality, in line with the predictive-selection account.

Next, we investigated whether the neural responses that were modulated by lexicality were also influenced by prime type by conducting an ROI analysis which tested the interaction between prime type and lexicality, as well as planned pairwise

**Table 4. Pre-DP MEG ROI analyses on all data versus data excluding items with pre-DP acoustic differences across gradiometer sensor locations that showed post-DP pseudoword > word effect<sup>a</sup>**

Contrast	All data		Data with exclusion	
	<i>t</i>	<i>p</i>	<i>t</i>	<i>p</i>
Unprimed > word-primed	2.41	0.013	2.57	0.009
Unprimed > pseudo-primed	2.69	0.014	3.14	0.005

<sup>a</sup>Reported effects on unprimed > word-primed items (planned) are one-tailed. Unprimed, unprimed items; Word-primed, word-primed items; pseudo-primed, pseudoword-primed items.

**Table 5. Post-DP MEG ROI analyses on all data versus data excluding items with pre-DP acoustic differences<sup>a</sup>**

Contrast	Modality	All data			Data with exclusion		
		<i>F</i>	<i>t</i>	<i>p</i>	<i>F</i>	<i>t</i>	<i>p</i>
Lexicality × prime type	Grad	7.30	—	0.002	6.12	—	0.005
	Mag	5.80	—	0.007	3.77	—	0.035
	Source	0.99	—	0.360	1.04	—	0.354
Word prime type	Grad	8.01	—	0.001	6.18	—	0.005
	Mag	5.61	—	0.009	4.46	—	0.021
	Source	3.77	—	0.038	3.64	—	0.039
Word-word > word	Grad	—	2.22	0.019	—	2.11	0.023
	Mag	—	3.33	0.002	—	2.79	0.006
	Source	—	2.66	0.007	—	2.51	0.010
Word-word > pseudo-word	Grad	—	3.70	<0.001	—	3.60	<0.001
	Mag	—	2.64	0.008	—	2.33	0.015
	Source	—	1.26	0.110	—	1.39	0.089
Pseudo prime type	Grad	0.67	—	0.514	0.57	—	0.564
	Mag	0.80	—	0.446	0.37	—	0.681
	Source	1.12	—	0.326	1.23	—	0.300

<sup>a</sup>Reported pairwise effects (planned) are one-tailed. Word-word, word-primed word; pseudo-word, pseudo-word-primed word; Grad, gradiometers; Mag, magnetometers.

comparisons of priming effects on words alone, using data averaged over the time window and the sensor locations of the significant cluster shown in Figure 6A, B (Fig. 6D,E; Table 5). Since these planned pairwise comparisons involve responses to familiar words only (i.e., words that are word-primed vs unprimed, words that are word-primed vs pseudoword-primed), they are orthogonal to the lexicality effect that defined the pseudoword > word cluster and hence are not confounded by task. The interaction was significant in both gradiometers,  $F_{(1.96,41.11)} = 7.30$ ,  $p = 0.002$ , and magnetometers,  $F_{(1.90,39.99)} = 5.80$ ,  $p = 0.007$ . Specifically, there was a significant effect of prime type for words,  $F_{(1.93,40.55)} = 8.01$ ,  $p = 0.001$  (gradiometers),  $F_{(1.81,37.96)} = 5.61$ ,  $p = 0.009$  (magnetometers), such that neural signals evoked by word-primed words were significantly stronger than those evoked by unprimed words,  $t_{(21)} = 2.22$ ,  $p = 0.019$  (gradiometers),  $t_{(21)} = 3.33$ ,  $p = 0.002$  (magnetometers), and pseudoword-primed words,  $t_{(21)} = 3.70$ ,  $p < 0.001$  (gradiometers),  $t_{(21)} = 2.64$ ,  $p = 0.008$  (magnetometers). In contrast, there was no reliable main effect of prime type for pseudowords,  $F_{(1.94,40.80)} = 0.67$ ,  $p = 0.514$  (gradiometers),  $F_{(1.79,37.61)} = 0.80$ ,  $p = 0.446$  (magnetometers).

The corresponding tests performed on the source-reconstructed power within the lexicality ROI of suprathreshold voxels (Fig. 6F; Table 5) did not show a reliable interaction effect between lexicality and competitor priming,  $F_{(1.56,32.85)} = 0.99$ ,  $p = 0.360$ . Nevertheless, consistent with sensor space results, source power indicated a significant effect of prime type for words,  $F_{(1.73,36.42)} = 3.77$ ,  $p = 0.038$ , but not pseudowords,  $F_{(1.62,33.94)} = 1.12$ ,  $p = 0.326$ . Pairwise comparisons also indicated that word-primed words evoked significantly greater source strength than unprimed words,  $t_{(21)} = 2.66$ ,  $p = 0.007$ , although the effect

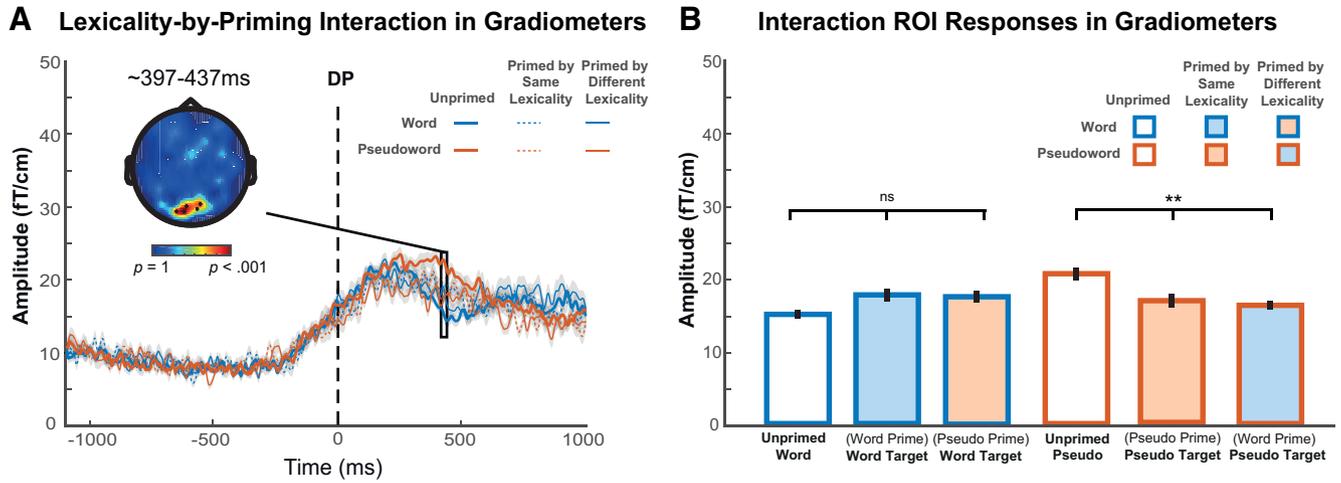
between word-primed and pseudoword-primed words was not significant,  $t_{(21)} = 1.26$ ,  $p = 0.110$ . Overall, in line with behavioral results, neural responses evoked by words and pseudowords were also influenced differently by prime type. Critically, competitor priming modulated the post-DP neural responses evoked by words, but not those evoked by pseudowords, and these effects were localized to the left STG regions that plausibly contribute to sublexical processing of speech. This matches the pattern of responses proposed in the predictive-selection model (see Fig. 1F). As shown in Tables 1–5, the pattern and significance of the results did not change when items with pre-DP acoustic differences identified through the gating post-test were excluded.

As encouraged by a reviewer, we also conducted whole brain analyses for the competitor priming effects. We found a significant word-primed word > unprimed word cluster of 1197 sensor × time points ( $p = 0.034$ ) in magnetometers in the left hemisphere within a time window of 426–466 ms post-DP. In addition, we also found a significant and a marginal word-primed word > pseudoword-primed word cluster in gradiometers in the left hemisphere, respectively, of 527 sensor × time points ( $p = 0.011$ ) at 719–749 ms and 471 sensor × time points ( $p = 0.053$ ) at 315–336 ms. These topographies and time courses overlap with the pseudoword > word clusters and are consistent with our ROI results. Hence, the ROI analyses have picked up the most important findings from these whole-brain analyses.

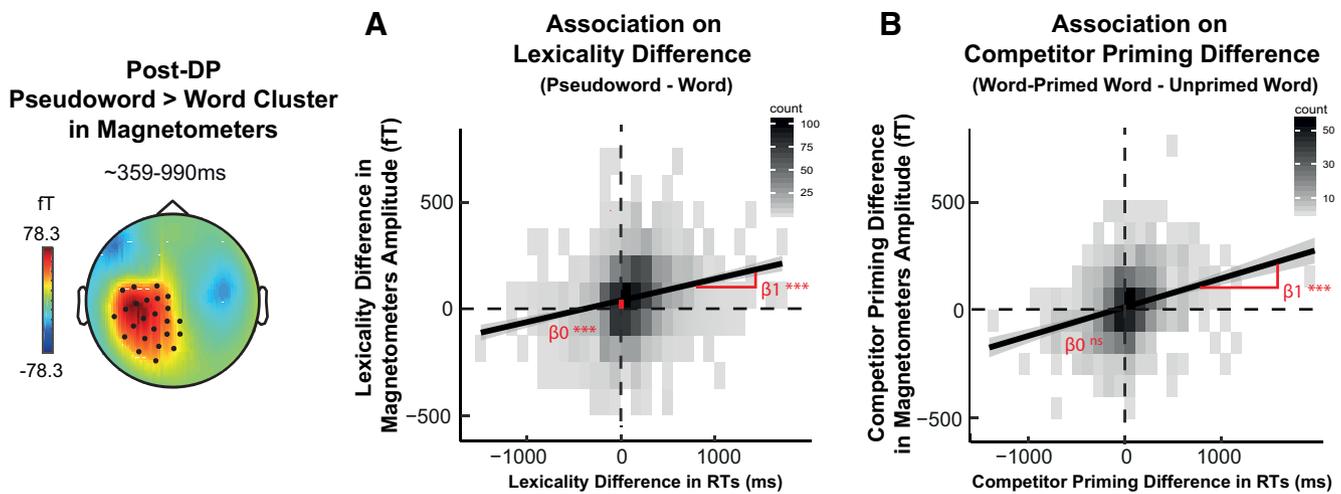
To ensure that other response patterns were not overlooked, we also investigated whether there was any lexicality by prime-type interaction at other locations across the scalp and source spaces, and during other time periods. As shown in Figure 7A, a significant cluster of gradiometers at midline posterior scalp locations were found at 397–437 ms post-DP, in which the effect of priming was significantly different for words and pseudowords. Figure 7B shows gradiometer signals evoked by conditions of interest averaged over the spatial and temporal extent of the significant cluster in Figure 7A. To explore this profile, we computed an orthogonal contrast to assess the overall lexicality effect (the difference between words and pseudowords), and the result was marginal,  $F_{(1.00,21.00)} = 3.50$ ,  $p = 0.075$ . The effect of prime type was marginally significant for words,  $F_{(1.89,39.78)} = 3.08$ ,  $p = 0.060$ , but significant for pseudowords,  $F_{(1.80,37.85)} = 7.14$ ,  $p = 0.003$ . The location and pattern of this interaction cluster were dissimilar to those predicted by either competitive- or predictive-selection theories, and no cluster survived correction in magnetometer sensors or source space; hence, we did not consider this effect to be as relevant or interpretable as our other findings. We report it here in the interest of completeness and transparency.

#### Linking neural and behavioral effects

To further examine the relationship between neural and behavioral response differences attributable to competitor priming or lexicality, we conducted a single-trial regression analyses using linear mixed-effect models that account for random intercepts and slopes for participants and stimuli sets (grouped by their initial segments). We calculated behavioral RT differences and neural MEG differences caused by: (1) lexicality, that is, the difference between pseudoword and word trials (collapsed over primed and unprimed conditions) and (2) competitor priming, that is, the difference between unprimed and word-primed word trials, with MEG signals averaged over the spatial and temporal extent of the post-DP pseudoword > word cluster seen in sensor space and the STG peak voxel in source space (Fig. 6). We then assessed the relationship between these behavioral and neural



**Figure 7.** Post-DP results showing lexicity × priming interaction effects in MEG gradiometers. **A**, The topographic plot represents *F* statistics for the statistically significant cluster that showed an interaction between lexicity and prime type. Waveforms represent gradiometer responses averaged over the spatial extent of the significant cluster shown in the topography. Gray shade of waveforms represents ± within-participant SE, adjusted to remove between-participants variance. **B**, Gradiometer signals evoked by conditions of interest averaged over temporal and spatial extent of the significant cluster in **A**. Error bars indicate ± within-participant SE, adjusted to remove between-participants variance. \*\**p* < 0.01. The statistical comparison lines indicate main effects of prime type for each lexicity. The lexicity × prime-type interaction is statistically reliable as expected based on the defined cluster.



**Figure 8.** Single-trial linear mixed-effect models, which accounted for random intercepts and slopes for participants and stimuli sets (grouped by initial segments), were constructed to compute the relationship between RTs and magnetometers on **(A)** lexicity difference (i.e., between pseudowords and words, collapsed over unprimed and primed conditions) and **(B)** competitor priming difference (i.e., between word-primed word and unprimed word conditions). Magnetometer responses were averaged over the time window and scalp locations of the significant post-DP pseudoword > word cluster (see Fig. 6).  $\beta_1$  refers to the model slope;  $\beta_0$  refers to the model intercept. \*\*\**p* < 0.001.

difference effects in linear mixed-effect regression of single trials, with differences in RTs as the independent variable and differences in MEG responses as the dependent variable. The analyses were conducted using the lme4 package in R (Bates et al., 2014).

As shown in Figure 8A, we observed a significant positive relationship between RTs and magnetometers on lexicity difference ( $\beta = 0.11$ ,  $SE = 0.01$ ,  $t_{(23.31)} = 7.77$ ,  $p < 0.001$ ), although associations between RTs and gradiometers or source response were not significant. These observations from magnetometers indicated that slower lexical decision times evoked by pseudowords were associated with greater neural responses. Furthermore, the intercept parameter for the magnetometers model was significantly larger than zero,  $\beta = 37.58$ ,  $SE = 5.72$ ,  $t_{(23.09)} = 6.57$ ,  $p < 0.001$ . We can interpret this intercept as the neural difference that would be predicted for trials in which there was no delayed response to pseudowords compared with words. The significant intercept indicated a baseline difference in neural responses to words and pseudowords,

even in the absence of any difference in processing effort (as indexed by lexical decision RTs). This suggested the engagement of additional neural processes specific to pseudowords regardless of the behavioral effect (compare Taylor et al., 2014).

Figure 8B showed another significant positive relationship between RTs and magnetometers on competitor priming difference ( $\beta = 0.15$ ,  $SE = 0.02$ ,  $t_{(38.85)} = 7.89$ ,  $p < 0.001$ ), while relationships between RTs and gradiometers or source response were again not significant. Interestingly, unlike for the lexicity effect, the intercept in this competitor priming magnetometers model did not reach significance ( $\beta = 12.88$ ,  $SE = 7.27$ ,  $t_{(21.33)} = 1.77$ ,  $p = 0.091$ ). This nonsignificant intercept might suggest that, if word-primed words did not evoke longer RTs than unprimed words, magnetometer signals would not be reliably different between the two conditions either. Hence, consistent with predictive-selection accounts, the increased post-DP neural responses in the STG caused by

competitor priming were both positively linked to and mediated by longer RTs.

## Discussion

In this study, we distinguished different implementations of Bayesian perceptual inference by manipulating the prior probability of spoken words and examining changes to neural responses. We replicated the competitor priming effect such that a single prior presentation of a competitor word (e.g., *hijack*) delayed the recognition of a similar-sounding word (e.g., *hygiene*), whereas this effect was not observed when the prime or target was a pseudoword (e.g., *hijure*). Armed with this behavioral evidence, we used MEG data to test the neural bases of two Bayesian theories of spoken word recognition.

### Competitive versus predictive selection

Competitive-selection accounts propose that word recognition is achieved through direct inhibitory connections between representations of similar candidates (e.g., McClelland and Elman, 1986). Priming boosts the activation of heard words and increases lateral inhibition applied to neighboring words, which delays their subsequent identification. The effect of competitor priming is to increase lexical uncertainty, and hence lexical-level neural responses, until later time points when target words can be distinguished from the competitor prime (Fig. 1C). In contrast, predictive-selection accounts propose that word recognition is achieved by subtracting predicted speech from heard speech and using computations of prediction error to update lexical probabilities (Davis and Sohoglu, 2020). By this view, predictions for segments that are shared between competitor primes and targets (pre-DP segments) will be enhanced after presentation of prime words. Thus, competitor priming will reduce the magnitude of prediction error, and hence neural responses pre-DP (Fig. 1F). Only when speech diverges from predictions (post-DP segments) will competitor-primed words evoke greater prediction error, leading to increased neural response in brain areas involved in prelexical (e.g., phonemic) processing of speech representing prediction error (Blank and Davis, 2016; Blank et al., 2018). Both these models involve multiple levels of representation and hence both sublexical and lexical processes. However, our focus is primarily on lexical processing within the competitive-selection framework and sublexical processing within the predictive-selection framework. These are the critical processing levels that (1) support word recognition, (2) are modulated by competitor priming effect, and (3) can potentially explain how slower behavioral responses will manifest in MEG responses.

The direction and timing of changes to MEG responses associated with competitor priming showed opposite effects pre- and post-DP. In the pre-DP period, consistent with predictive-selection but contrary to competitive-selection mechanisms, we saw decreased neural responses for word-primed items compared with unprimed items. The initial, shared segments between prime (*hijack*) and target (*hygiene*) words evoked a reduced response during early time periods in line with a reduction in prediction error. However, during the post-DP period, competitor-primed words evoked stronger neural responses than unprimed words in exactly the same locations and time periods that showed increased responses to pseudowords (*hijure*) compared with words. These post-DP response increases are in line with enhanced processing difficulty for competitor-primed words and pseudowords because of greater prediction error.

Thus, the time course of the competitor priming neural effects, showing reduced neural responses pre-DP and increased neural responses post-DP, closely resembles the expected changes in prediction error (Fig. 1F) based on predictive-selection mechanisms. However, we note that post-DP effects reach significance later than mismatch effects for written words (Dikker et al., 2010), lexicality effects for spoken words (MacGregor et al., 2012), and phoneme surprisal effects in connected speech (Brodbeck et al., 2018; Donhauser and Baillet, 2020). This delay could be because of our morphing manipulation which removed coarticulation before the divergence point, or because of neural effects being delayed for words in isolation compared with connected speech (for review, see Gwilliams and Davis, 2021). Further research to assess the latency of neural effects can help determine whether they are sufficiently early to indicate bottom-up sensory signals as proposed by predictive selection.

Effects of lexicality and competitor priming localized to the left STG; this brain region has long been associated with lower-level sensory processing of speech (Yi et al., 2019). Our observation of increased responses to pseudowords in STG agrees with source-localized MEG findings (Gagnepain et al., 2012; Shtyrov et al., 2012) and a meta-analysis of PET and fMRI studies (Davis and Gaskell, 2009). This location is also consistent with the proposal that lexical influences on segment-level computations produce reliable neural differences between words and pseudowords (Davis and Sohoglu, 2020). We take this localization as further evidence in favor of computations of segment prediction error as a critical mechanism underlying word identification.

We further show using regression analyses that neural (MEG) and behavioral (RT) effects of lexicality and competitor priming are linked on a trial-by-trial basis. Trials in which pseudoword processing or competitor priming leads to larger increases in RT also have greater post-DP neural responses. Links between behavioral and neural effects of lexicality and competitor priming are once more in line with the proposal that post-DP increases in prediction error are a key neural mechanism for word and pseudoword processing and explain delayed behavioral responses seen in competitor priming. Interestingly, lexicality and competitor priming effects differ in terms of whether a reliable neural response difference would be seen for trials with no baseline RT difference. While neural lexicality effects were significant, even for trials that did not show behavioral effects, the same was not true for the competitor priming effect. These results indicate that, consistent with predictive-selection accounts, the post-DP neural effect of competitor priming was mediated by changes in behavioral RTs. In contrast, an increased neural response to pseudowords was expected even in trials for which RTs did not differ between pseudowords and words. We next consider the implications of these and other findings for pseudoword processing.

### How do listeners process pseudowords?

Participants identified pseudowords with a speed and accuracy similar to that seen during recognition of familiar words. This is consistent with an optimally efficient language processing system (Marslen-Wilson, 1984; Zhuang et al., 2014), in which pseudowords can be distinguished from real words as soon as deviating speech segments are heard. Beyond this well-established behavioral finding, however, we reported two seemingly contradictory observations concerning pseudoword processing.

The first is that, while post-DP neural activity and RTs for words were modulated by competitor priming, processing of pseudowords was not similarly affected. This might suggest that

the prior probability of hearing a pseudoword and the prediction error elicited by mismatching segments are not changed by our experimental manipulations. This may be because pseudowords have a low or zero prior probability and elicit maximal prediction errors that cannot be modified by a single prime. Yet, memory studies suggest that even a single presentation of a pseudoword can be sufficient for listeners to establish a lasting memory trace (Mckone and Trynes, 1999; Arndt et al., 2008). However, it is possible that this memory for pseudowords reflects a different type of memory (e.g., episodic memory) from that produced by a word, with only the latter able to temporarily modify long-term, lexical-level representations and predictions for word speech segments (as in Complementary Learning Systems theories, McClelland et al., 1995; Davis and Gaskell, 2009).

A second observation is that, contrary to the null result for post-DP processing, pseudoword priming reduced subsequent pre-DP neural responses evoked by target items to a similar degree as word priming (Fig. 5B). This pre-DP effect is surprising given previous evidence suggesting that pseudowords must be encoded into memory and subject to overnight, sleep-associated consolidation to modulate the speed of lexical processing (Tamminen et al., 2010; James et al., 2017) or neural responses (Davis and Gaskell, 2009; Landi et al., 2018). It might be that neural effects seen for these pre-DP segments were because of changes to the representation of familiar words that our pseudowords resembled, although these were insufficient to modulate processing of post-DP segments.

In conclusion, our work provides compelling evidence in favor of neural computations of prediction error during spoken word recognition. Although previous work by Gagnepain et al. (2012) provided evidence for the predictive-selection account, their behavioral effects of consolidation on word recognition were obtained during different tasks and different test sessions from neural responses. Our current study goes beyond this previous work by adopting a single task (lexical decision) and using a competitor priming paradigm that permits concurrent measurement of perceptual outcomes and neural responses in a single session. This enables us to directly link trials that evoked stronger neural signals in the STG to delayed RTs and hence provide stronger evidence that both of these effects are caused by competitor priming.

In addition, unlike previous work (Brodbeck et al., 2018; Donhauser and Baillet, 2020), which reported neural responses correlated with lexical entropy as well as prediction error (surprisal), we did not find similarly equivocal evidence. These earlier studies measured neural responses to familiar words in continuous speech sequences, such as stories or talks. It might be that effects of lexical entropy are more apparent for connected speech than isolated words. However, since lexical uncertainty (entropy) and segment-level predictability (segment prediction error or surprisal) are highly correlated in natural continuous speech, these studies may be less able to distinguish between the lexical and segmental mechanisms that we assessed here. In contrast, our speech materials were carefully selected to change lexical probability (through priming) and for priming to have opposite effects on segment prediction error before and after DP. This manipulation provides evidence in favor of predictive-selection mechanisms that operate using computations of prediction error during spoken word recognition.

## References

- Aitchison L, Lengyel M (2017) With or without you: predictive coding and Bayesian inference in the brain. *Curr Opin Neurobiol* 46:219–227.
- Arndt J, Lee K, Flora DB (2008) Recognition without identification for words, pseudowords and nonwords. *J Mem Lang* 59:346–360.
- Baayen RH, Piepenbrock R, Van Rijn H (1993) The CELEX lexical database (CD-ROM). Linguistic data consortium. Philadelphia, PA: University of Pennsylvania.
- Barr DJ, Levy R, Scheepers C, Tily HJ (2013) Random effects structure for confirmatory hypothesis testing: keep it maximal. *J Mem Lang* 68:255–278.
- Bates D, Mächler M, Bolker B, Walker S (2014) Fitting linear mixed-effects models using lme4. arXiv 1406.
- Blank H, Davis MH (2016) Prediction errors but not sharpened signals simulate multivoxel fMRI patterns during speech perception. *PLoS Biol* 14: e1002577.
- Blank H, Spangenberg M, Davis MH (2018) Neural prediction errors distinguish perception and misperception of speech. *J Neurosci* 38:6076–6089.
- Bozic M, Tyler LK, Ives DT, Randall B, Marslen-Wilson WD (2010) Bihemispheric foundations for human speech comprehension. *Proc Natl Acad Sci USA* 107:17439–17444.
- Brodbeck C, Hong LE, Simon JZ (2018) Rapid transformation from auditory to linguistic representations of continuous speech. *Curr Biol* 28:3976–3983.
- Brysbaert M, Stevens M, Mander P, Keuleers E (2016) How many words do we know? Practical estimates of vocabulary size dependent on word definition, the degree of language input and the participant's age. *Front Psychol* 7:1116–1111.
- Bürkner PC (2017) brms: an R package for Bayesian multilevel models using Stan. *J Stat Softw* 80:1–28.
- Cousineau D (2005) Confidence intervals in within-participant designs: a simpler solution to Loftus and Masson's method. *TQMP* 1:42–45.
- Davis MH, Gaskell MG (2009) A complementary systems account of word learning: neural and behavioural evidence. *Philos Trans R Soc Lond B Biol Sci* 364:3773–3800.
- Davis MH, Scharenborg O (2016) Speech perception by humans and machines. In: Gaskell G, Mirkovic J (eds). *Speech perception and spoken word recognition*. London: Psychology.
- Davis MH, Sohoglu E (2020) Three functions of prediction error for Bayesian inference in speech perception. In: Gazzaniga M, Mangun R, Poeppel D (eds), *The cognitive neurosciences*, Ed 6. Cambridge, MA: Massachusetts Institute of Technology.
- Davis MH, Marslen-Wilson WD, Gaskell MG (2002) Leading up the lexical garden path: segmentation and ambiguity in spoken word recognition. *J Exp Psychol Hum Percept Perform* 28:218.
- de Leeuw JR (2015) jsPsych: a JavaScript library for creating behavioral experiments in a web browser. *Behav Res Methods* 47:1–12.
- Dikker S, Rabagliati H, Farmer TA, Pylkkänen L (2010) Early occipital sensitivity to syntactic category is based on form typicality. *Psychol Sci* 21:629–634.
- Donhauser PW, Baillet S (2020) Two distinct neural timescales for predictive speech processing. *Neuron* 105:385–393.e9.
- Gagnepain P, Henson RN, Davis MH (2012) Temporal predictive codes for spoken words in auditory cortex. *Curr Biol* 22:615–621.
- Gwilliams LE, Davis MH (2021) Extracting language content from speech sounds: the information theoretic approach. In: *The Auditory cognitive neuroscience of speech perception: Springer handbook of auditory research* (Holt LL, Peelle JE, Coffin AB, Popper AN, Fay RR eds). New York: Springer.
- Henson RN, Wakeman DG, Litvak V, Friston KJ (2011) A Parametric Empirical Bayesian framework for the EEG/MEG inverse problem: generative models for multiparticipant and multimodal integration. *Front Hum Neurosci* 5:1–16.
- James E, Gaskell MG, Weighall A, Henderson L (2017) Consolidation of vocabulary during sleep: the rich get richer? *Neurosci Biobehav Rev* 77:1–13.
- Kawahara H (2006) STRAIGHT, exploitation of the other aspect of VOCODER: perceptually isomorphic decomposition of speech sounds. *Acoust Sci Tech* 27:349–353.
- Kilner JM, Friston KJ (2010) Topological inference for EEG and MEG. *Ann Appl Stat* 4:1272–1290.
- Kisler T, Reichel UD, Schiel F (2017) Multilingual processing of speech via web services. *Comput Speech Lang* 45:326–347.
- Kleiner M, Brainard D, Pelli D, Ingling A, Murray R, Broussard C (2007) What's new in Psychtoolbox-3. *Perception* 36:1–16.

- Landi N, Malins JG, Frost SJ, Magnuson JS, Molfese P, Ryherd K, Rueckl JG, Mencl WE, Pugh KR (2018) Neural representations for newly learned words are modulated by overnight consolidation, reading skill, and age. *Neuropsychologia* 111:133–144.
- Litvak V, Mattout J, Kiebel S, Phillips C, Henson R, Kilner J, Barnes G, Oostenveld R, Daunizeau J, Flandin G, Penny W, Friston K (2011) EEG and MEG data analysis in SPM8. *Comput Intell Neurosci* 2011:852961.
- MacGregor LJ, Pulvermüller F, Van Casteren M, Shtyrov Y (2012) Ultra-rapid access to words in the brain. *Nat Commun* 3:1–7.
- Mandera P, Keuleers E, Brysbaert M (2017) Explaining human performance in psycholinguistic tasks with models of semantic similarity based on prediction and counting: a review and empirical validation. *J Mem Lang* 92:57–78.
- Marslen-Wilson WD (1984) Function and process in spoken word recognition. In: Bouma H, Bouwhuis DG (eds), pp 125–150. *Attention and performance X: control of language processes*. Hillsdale, NJ: Erlbaum.
- Marsolek CJ (2008) What antipriming reveals about priming. *Trends Cogn Sci* 12:176–181.
- Mattout J, Henson RN, Friston KJ (2007) Canonical source reconstruction for MEG. *Comput Int Neurosci* 2007:67613.
- McClelland JL, Elman JL (1986) The TRACE model of speech perception. *Cogn Psychol* 18:1–86.
- McClelland JL, McNaughton BL, O'Reilly RC (1995) Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychol Rev* 102:419–457.
- Mckone E, Trynes K (1999) Acquisition of novel traces in short-term implicit memory: priming for nonwords and new associations. *Mem Cognit* 27:619–632.
- Monsell S, Hirsh KW (1998) Competitor priming in spoken word recognition. *J Exp Psychol Learn Mem Cogn* 24:1495–1520.
- Norris D, McQueen JM (2008) Shortlist B: a Bayesian model of continuous speech recognition. *Psychol Rev* 115:357–395.
- Norris D, McQueen JM, Cutler A (2002) Bias effects in facilitatory phonological priming. *Mem Cognit* 30:399–411.
- Okada K, Hickok G (2006) Identification of lexical–phonological networks in the superior temporal sulcus using functional magnetic resonance imaging. *Neuroreport* 17:1293–1296.
- Prabhakaran R, Blumstein SE, Myers EB, Hutchison E, Britton B (2006) An event-related fMRI investigation of phonological–lexical competition. *Neuropsychologia* 44:2209–2221.
- Rogers JC, Davis MH (2017) Inferior frontal cortex contributions to the recognition of spoken words and their constituent speech sounds. *Journal of cognitive neuroscience* 29:919–936.
- Schiel F (1999) Automatic phonetic transcription of non-prompted speech. *Proceedings of the 14th International Congress of Phonetic Sciences*, pp 607–610.
- Shtyrov Y, Smith ML, Horner AJ, Henson R, Nathan PJ, Bullmore ET, Pulvermüller F (2012) Attention to language: novel MEG paradigm for registering involuntary language processing in the brain. *Neuropsychologia* 50:2605–2616.
- Sohoglu E, Davis MH (2016) Perceptual learning of degraded speech by minimizing prediction error. *Proc Natl Acad Sci USA* 113:E1747–E1756.
- Sohoglu E, Peelle JE, Carlyon RP, Davis MH (2012) Predictive top-down integration of prior knowledge during speech perception. *J Neurosci* 32:8443–8453.
- Spratling MW (2008) Reconciling predictive coding and biased competition models of cortical function. *Front Comput Neurosci* 2:4.
- Tamminen J, Payne JD, Stickgold R, Wamsley EJ, Gaskell MG (2010) Sleep spindle activity is associated with the integration of new memories and existing knowledge. *J Neurosci* 30:14356–14360.
- Tauroza S, Allison D (1990) Speech rates in British English. *Appl Linguistics* 11:90–105.
- van Casteren M, Davis MH (2006) Mix, a program for pseudorandomization. *Behav Res Methods* 38:584–589.
- Taylor JS, Rastle K, Davis MH (2014) Interpreting response time effects in functional imaging studies. *Neuroimage* 99:419–433.
- van Heuven WJ, Mandera P, Keuleers E, Brysbaert M (2014) SUBTLEX-UK: a new and improved word frequency database for British English. *Q J Exp Psychol* 67:1176–1190.
- Wagenmakers EJ, Lodewyckx T, Kuriyal H, Grasman R (2010) Bayesian hypothesis testing for psychologists: a tutorial on the Savage–Dickey method. *Cogn Psychol* 60:158–189.
- Yi HG, Leonard MK, Chang EF (2019) The encoding of speech sounds in the superior temporal gyrus. *Neuron* 102:1096–1110.
- Zhuang J, Randall B, Stamatakis EA, Marslen-Wilson WD, Tyler LK (2011) The interaction of lexical semantics and cohort competition in spoken word recognition: an fMRI study. *J Cogn Neurosci* 23:3778–3790.
- Zhuang J, Tyler LK, Randall B, Stamatakis EA, Marslen-Wilson WD (2014) Optimally efficient neural systems for processing spoken language. *Cereb Cortex* 24:908–918.