Systems/Circuits

Parametric Cognitive Load Reveals Hidden Costs in the Neural Processing of Perfectly Intelligible Degraded Speech

Harrison Ritz,[1,2] **Conor J. Wild,**[1] **and Ingrid S. Johnsrude**[1,3]

[1]Brain and Mind Institute, University of Western Ontario, London, Ontario N6A 3K7, Canada, [2]Department of Cognitive, Linguistic, and Psychological Sciences, Brown University, Providence, Rhode Island 02912, and [3]Departments of Psychology and Communication Sciences and Disorders, University of Western Ontario, London, Ontario N6A 3K7, Canada

Speech is often degraded by environmental noise or hearing impairment. People can compensate for degradation, but this requires cognitive effort. Previous research has identified frontotemporal networks involved in effortful perception, but materials in these works were also less intelligible, and so it is not clear whether activity reflected effort or intelligibility differences. We used functional magnetic resonance imaging to assess the degree to which spoken sentences were processed under distraction and whether this depended on speech quality even when intelligibility of degraded speech was matched to that of clear speech (close to 100%). On each trial, male and female human participants either attended to a sentence or to a concurrent multiple object tracking (MOT) task that imposed parametric cognitive load. Activity in bilateral anterior insula reflected task demands; during the MOT task, activity increased as cognitive load increased, and during speech listening, activity increased as speech became more degraded. In marked contrast, activity in bilateral anterior temporal cortex was speech selective and gated by attention when speech was degraded. In this region, performance of the MOT task with a trivial load blocked processing of degraded speech, whereas processing of clear speech was unaffected. As load increased, responses to clear speech in these areas declined, consistent with reduced capacity to process it. This result dissociates cognitive control from speech processing; substantially less cognitive control is required to process clear speech than is required to understand even very mildly degraded, 100% intelligible speech. Perceptual and control systems clearly interact dynamically during real-world speech comprehension.

Key words: **cognitive control; functional magnetic resonance imaging; speech perception**

Significance Statement

Speech is often perfectly intelligible even when degraded, for example, by background sound, phone transmission, or hearing loss. How does degradation alter cognitive demands? Here, we use fMRI to demonstrate a novel and critical role for cognitive control in the processing of mildly degraded but perfectly intelligible speech. We compare speech that is matched for intelligibility but differs in putative control demands, dissociating cognitive control from speech processing. We also impose a parametric cognitive load during perception, dissociating processes that depend on tasks from those that depend on available capacity. Our findings distinguish between frontal and temporal contributions to speech perception and reveal a hidden cost to processing mildly degraded speech, underscoring the importance of cognitive control for everyday speech comprehension.

Introduction

In perfect listening conditions, the comprehension of speech is seemingly effortless for healthy young people. However, everyday listening conditions are rarely as good as in the laboratory, and speech understanding is often compromised by noisy environments, low-fidelity digital communication, and hearing impairment. Listeners must exert cognitive control to understand markedly degraded speech (Broadbent, 1958; Vaden et al., 2013; Fedorenko, 2014; Heald and Nusbaum, 2014; Eckert et al., 2016; Johnsrude and Rodd, 2016; Pichora-Fuller et al., 2016; Rouault and Koechlin, 2018). However, what about very mildly degraded, perfectly intelligible speech? Does this also require attention and cognitive control, and if so, how much? A powerful method for quantifying control demands is to measure how processing of speech changes with declining speech quality and under distraction. Neuroimaging experiments have revealed that cingulo-opercular regions associated with cognitive control (Shenhav et al., 2013) and temporal regions associated with high-level speech perception (Hickok and Poeppel, 2007) are sensitive to speech intelligibility (Davis and Johnsrude, 2003; Eckert et al., 2016), lose

Received Sep. 1, 2021; revised Mar. 8, 2022; accepted Mar. 10, 2022.

Author contributions: H.R. and I.S.J. designed research; H.R. and C.J.W. performed research; H.R. and C.J.W. analyzed data; H.R. and I.S.J. wrote the paper.

This work was supported by a Canadian Institutes of Health Research Operating Grant and a Natural Sciences and Engineering Research Council of Canada Discovery Grant to I.S.J. We thank Natalie R. Osborne for assistance with data collection.

The authors declare no competing financial interests.

Correspondence should be addressed to Harrison Ritz at harrison.ritz@gmail.com.

https://doi.org/10.1523/JNEUROSCI.1777-21.2022

speech sensitivity during distracting tasks (Sabri et al., 2008; Wild et al., 2012), and predict perceptual accuracy perceptual accuracy (Wild et al., 2012; Vaden et al., 2013, 2015, 2016).

The existing body of research generally supports a role for domain-general control networks in degraded speech. However, this work has been limited in its ability to parcellate regions into those that are speech selective and those that respond in a domain-general fashion to all task demands. In a previous neuroimaging experiment, we found a set of frontal and temporal regions in which activity correlated with intelligibility when participants attended to speech, but not when they attended to either visual or auditory distractor tasks (Wild et al., 2012). In this study, clear and degraded speech was not matched on intelligibility, limiting our ability to dissociate general and specific contributions to speech perception. For example, a domain-general region that monitors or controls task performance could appear sensitive to speech intelligibility during comprehension tasks, but only because intelligibility is strongly correlated with accuracy. In contrast, responses in a domain-specific region involved in effortful speech perception should distinguish between clear and intelligibility-matched degraded speech when control capacity is sufficiently taxed to disrupt degraded speech processing. These two functions are likely to be organized hierarchically, with domain-general control processes in inferior frontal regions and speech-selective processing in temporal regions of the frontotemporal language processing system (Davis and Johnsrude, 2003; Hickok and Poeppel, 2007; Evans and Davis, 2015). In the current study, we compare perception of clearly spoken sentences with perception of sentences matched for intelligibility (near-perfect word report accuracy) and sentences with only slightly lower intelligibility (>90% word report accuracy), allowing us to dissociate intelligibility from putative control demands.

As in our previous experiment (Wild et al., 2012), we measured speech processing when listeners either attended to the speech or when it was presented, while listeners were performing a distracting task. Critically, the stimuli across all conditions were identical (over participants); all that differed were participants' task goals. This allows us to isolate the influences of top-down attention. To better understand the trade-offs in resource allocation between these two concurrent tasks, we parametrically varied cognitive load and compared BOLD responses to intelligibility-matched clear and degraded speech under these different levels. This novel parametric manipulation distinguishes processes that depend on whether speech is relevant for the current task (task-dependent control; Speech × Task interaction) from processes that depend on the amount of control capacity that is available to aid perception (load-dependent control; Speech × Task × Load interaction). This parametric approach can help identify regions subserving domain-general processes (e.g., monitoring of accuracy, regardless of task) while also clarifying the role of control in domain-specific speech processes (e.g., identifying when speech-sensitive regions have a parametric versus all-or-nothing dependence on control).

We demonstrate that the focus of attention, whether individuals were listening to speech or doing multiple object tracking, had a strikingly different effect on neural response to clear and degraded speech in high-level speech regions. Whereas the responses in anterior insulae were consistent with domain-general performance monitoring, anterior temporal cortex was selectively recruited for speech perception, with a strikingly different response profile for clear and intelligibility-matched degraded speech under parametric cognitive load. These results reveal the division of labor within a classical frontotemporal speech network, where cognitive control enhances speech perception in challenging listening conditions.

Materials and Methods

Participants. Twenty-six individuals (15 females; M_{age} = 21.5, SD_{age} = 3.86) participated in this experiment after providing informed consent in accordance with the research ethics board at the University of Western Ontario. Participants were right-handed, native English speakers, with self-reported normal (or corrected to normal) vision and self-reported normal hearing. Two participants were removed before analysis because of dislodged earbuds and excessive movement during scanning, leaving 24 participants for the subsequent analyses. We chose our sample size to be at least as large as our previous fMRI experiment (n = 21 in Wild et al., 2012), which found strong interactions between speech type and task in cortical BOLD responses [e.g., Clear speech vs six-band noise vocoded (NV6) speech crossed with attend-speech vs attend-vision: interaction d = 1.6 in superior temporal sulcus, d = 0.92 in left inferior frontal gyrus].

Experimental design. On every trial, participants both heard a sentence and saw moving dots (Fig. 1). At the beginning of each trial, we instructed participants to either attend to the speech (LISTEN) or to perform a visual tracking task (TRACK). Across trials, we manipulated which task participants performed (two levels), the clarity of speech that participants heard (three levels), and the number of dots that participants saw on their screen (four levels), generating 24 factorial conditions. Participants experienced three trials from each condition in each of the three scanning runs, for a total of 216 experimental trials. Participants also experienced two types of control trial—24 silent, fixation-only trials and 24 LISTEN trials with rotated NV speech (see below), distributed equally across the three runs. We block randomized conditions within each scanner run to minimize the effect of low-frequency drift.

Speech task (LISTEN). Because of a technical error, the comprehension and tracking data during scanning were lost for two participants, leaving 22 participants for behavioral analyses.

We used the same materials as in a previous experiment (Wild et al., 2012), 216 everyday sentences, all recorded by the same female speaker of Canadian English (e.g., "His handwriting was very difficult to read"). Stimuli were presented diotically via foam-tipped insert earphones (Sensimetrics) at a comfortable listening level. The sentences were 6–13 words long, 1.2–4.7 s in duration, and split into six lists that were closely matched on the number of words, the sentence duration, and the summed word frequency (Thorndike–Lorge written frequency). These lists were assigned to the six Speech × Task conditions, counterbalanced across participants.

The clarity of the speech stimuli was manipulated using noise vocoding (Shannon et al., 1995). Each speech signal was filtered into logarithmically spaced frequency bands, with boundaries chosen to be equally spaced along the basilar membrane (Greenwood, 1990). The amplitude envelope within each frequency band was extracted and convolved with white noise that was band limited to the same frequency range. Previous work has found that intelligibility depends on the number of bands (Shannon et al., 1995; Davis and Johnsrude, 2003). In this experiment, we used highly intelligible noise-vocoded stimuli, filtered with 12 (NV12) and 6 (NV6) bands, as well as Clear (unmanipulated) speech. Unintelligible, spectrotemporally matched control stimuli were generated by spectral rotation; during the vocoding process, we permuted the assignment of speech envelopes to their noise envelopes (i.e., randomized over frequency bands; Blesser, 1972).

The usual measure of intelligibility is the number of words from an utterance that a listener can report correctly (word report accuracy). We chose a two-alternative gist-report measure because it is well matched to responses to our multiple object tracking (MOT) task and avoids the motion artifacts that word report may produce. To validate our gist-based measure of intelligibility, we drew from a pilot experiment using a word-report measure (Wild et al., 2012). In this experiment, a different group of participants heard the materials used in our experiment (sentences presented as Clear, NV12, and NV6), and their immediate verbal recall was scored for the percentage of words accurately reported. This

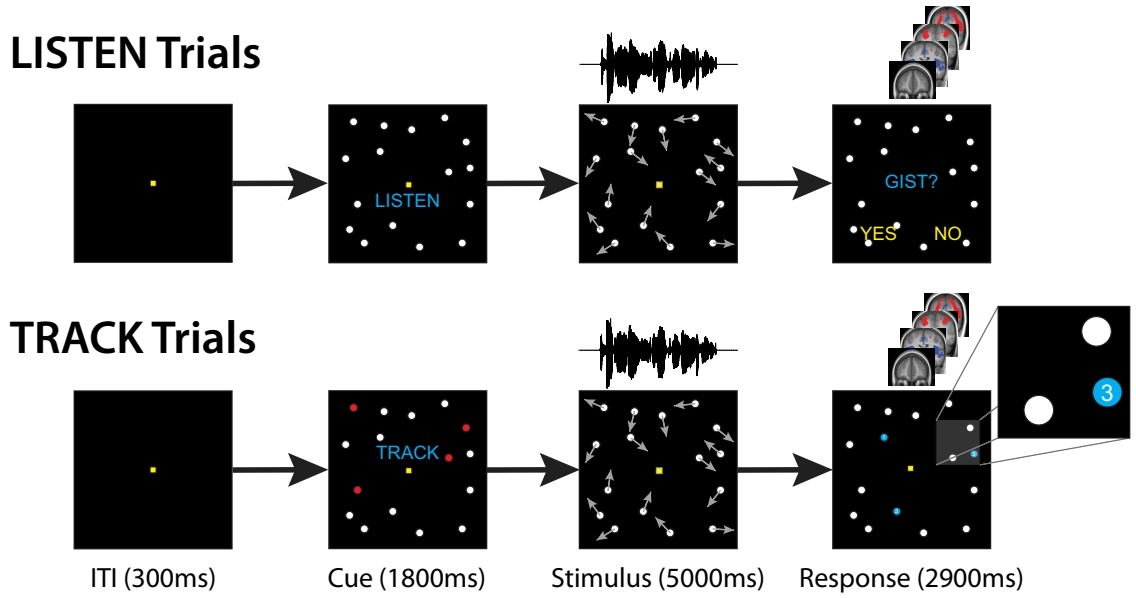

Figure 1. Trial time course. At the beginning of each trial, participants were first cued to focus on speech (LISTEN) or focus on tracking (TRACK). They then both heard speech and saw moving dots, making a response during the whole-brain fMRI acquisition (occurring 4 s after stimulus midpoint). Speech stimuli were ordinary sentences (e.g., "Her handwriting was very difficult to read") that were either clear (undistorted), 12-band noise vocoded, or 6-band noise vocoded, and during LISTEN trials participants reported whether they understood the gist of each sentence. During tracking, participants tracked 1, 3, 4, or 6 moving dots among 12 distractors and then reported which queried dot had been a member of the tracked set.

work demonstrated that gist-based measures are highly consistent with word report scores (Fig. 2*A*; Wild et al., 2012; Fig. 2).

Volumes were collected using a sparse acquisition protocol (Hall et al., 1999) in which our speech stimuli were presented during the silent period (9 s) between scans. The onset of each scan began 4 s after the midpoint of each sentence and tracking task, sampling the hemodynamic response near its peak amplitude. On LISTEN trials, participants had 2.8 s near the end of the 9 s silent period to indicate with a yes/no keypress (dominant hand) whether they had understood the gist of the sentence (Fig. 1). We analyzed the gist report rate using a logistic mixed-effects regression [MATLAB fitglme function; gist ~1 + speechType + (1 + speechType | participant)], using a maximal random effects structure throughout (Barr et al., 2013).

Multiple object tracking task (TRACK). Between 13 and 18 dots were on the screen throughout every trial, regardless of the task. All dots had a diameter of ~1° of visual angle and were shown against a black screen spanning ~20 × 20°. Dots were stationary for 1.8 s and then moved pseudorandomly around the screen at an approximate speed of 1.8° per second, with dots repelling 180° away from other dots or the edge of the screen at a 0.5° proximity.

On TRACK trials, participants tracked a subset of the moving dots (MOT; Pylyshyn and Storm, 1988). On these trials, 1, 3, 4, or 6 target dots were highlighted in red for 1.8 s before movement. Participants were instructed to keep their gaze on a fixation cue in the center of the screen and track these dots covertly. After 5 s of tracking, the dots froze in place, and three dots (one randomly selected target and two foils) were highlighted in blue and labeled 1, 2, and 3. Participants had 2.8 s to indicate with a three-alternative keypress which of the numbered dots was a target, without feedback (Fig. 1). We analyzed tracking accuracy using a logistic mixed-effects regression [MATLAB fitglme function; accuracy ~1 + load + (1+ load | participants)], and we analyzed accurate log reaction times using linear mixed-effects regression [MATLAB fitlme function; log(RT) ~ 1 + load + (1 + load | participants)].

Pretraining and memory post test. Before the scanning session, participants separately practiced both the speech and tracking tasks. First, participants were familiarized with NV speech to bring their comprehension performance to asymptote (Davis et al., 2005). Over 24 trials, participants heard a noise-vocoded sentence, indicated whether they had understood the gist of the sentence, and then received feedback by hearing the vocoded sentence again while also reading it on the screen (following the most effective training protocol according to Davis et al., 2005, experiment 3). In a second task, participants practiced MOT over 24 trials. On the first 12 trials, the number of targets began at one and increased (to three, four, and six) after each correct tracking response, or it decreased after each incorrect response. On the last 12 trials, the number of targets on each trial was randomly selected (from one, three, four, or six).

After the scanning session, we tested participants on their recognition memory for the sentences they had heard. On each trial, participants saw a written sentence on a computer screen and indicated with a keypress whether they remembered this sentence from the experiment (OLD) or whether it was new (NEW). Participants were tested on all 216 sentences from the experiment, along with 108 foil sentences. Foil sentences differed from target sentences in their topic, content words, and linguistic properties, but because target sentences were counterbalanced across conditions, foil properties could not systematically bias recognition memory, and we assess relative change in recognition performance across conditions, which is unaffected by foil type. During the scanning session, participants were unaware that memory would be tested, ensuring that memory encoding was incidental. To provide a signal-detection analysis (DeCarlo, 1998), we analyzed participants' recognition accuracy using probit mixed-effects regression [MATLAB fitglme function; correctRecognition ~1 + falseAlarmRate + speechType * task + (1 + speechType * task | participant)].

fMRI acquisition. Images were acquired on the 3.0T Siemens Prisma MRI system at the University of Western Ontario. T1-weighted structural images were collected at the beginning of each session using a single-shot EPI (FOV, 256 mm^2; resolution, 1 mm isotropic; slice thickness, 1 mm with 50% gap; TE, 2.98 ms; TR, 2300 ms; flip angle, 9°). T2*-weighted functional volumes were acquired across the whole brain using a four-factor interleaved multiband gradient EPI (FOV, 192 mm^2; resolution: 2.5 mm isotropic; slice thickness: 2.5 mm with 10% gap; 52 slices; TE: 30 ms; TA: 1000 ms; TR: 10 s; flip angle: 70°). Acquisition was transverse oblique, angled away from the eyes.

fMRI preprocessing and analysis. fMRI data were preprocessed and analyzed using SPM12 (Wellcome Center for Neuroimaging), following standard preprocessing steps including realignment, coregistration, and simultaneous segmentation and normalization to Montreal Neurological Institute (ICBM452) space. Normalization parameters were calculated from the structural image and applied to functional images coregistered to the mean of each run, resampling the images at 2 mm^3. The normalized images were spatially smoothed using a 3D Gaussian kernel with an 8 mm FWHM.

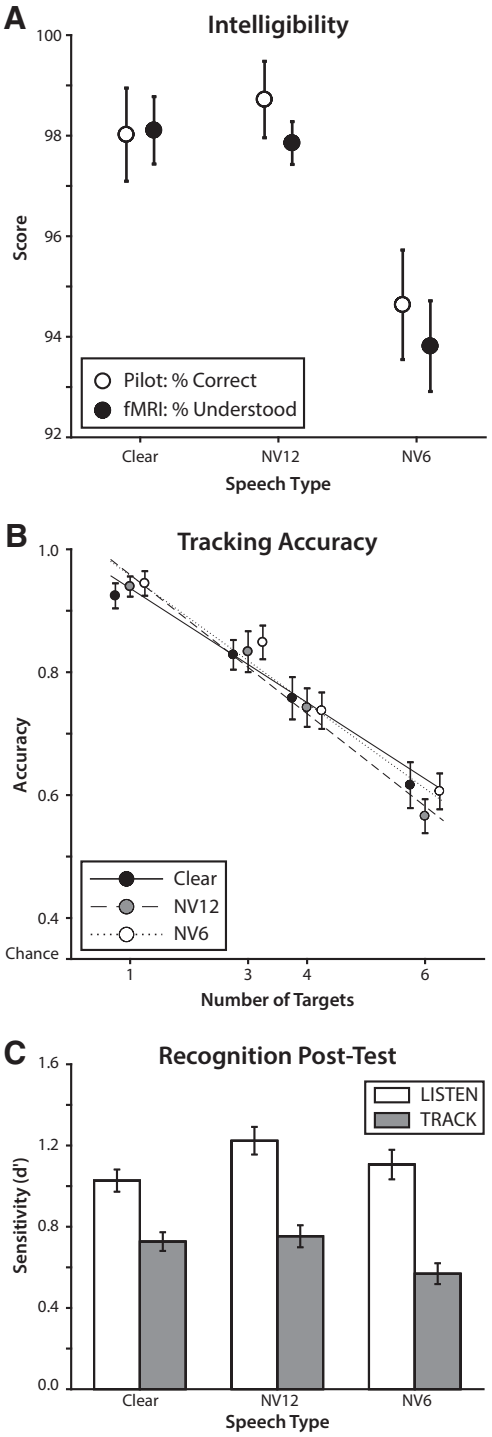

Figure 2. Behavioral results. ***A***, Intelligibility. Intelligibility scores across speech types were similar whether measured as objective word report accuracy (behavioral pilot; $n = 12$) or as subjective gist report (scanner experiment; $n = 22$). ***B***, Tracking accuracy. When participants tracked more targets, their tracking accuracy declined. Participants' accuracy remained above chance (33%) at all levels of tracking load. ***C***, Recognition post test. After the main experiment, participants performed a surprise memory test for the speech stimuli, deciding whether written probes had been heard previously or were novel. Memory sensitivity was quantified with *d'*, comparing hit and false alarm rates. All error bars indicate within-participant SEM (Morey, 2008).

Statistical parametric maps for each subject were estimated using a general linear model containing onset indicators for rotated speech and the six combinations of Speech (Clear, NV6, and NV12) by Task (LISTEN and TRACK) conditions. The model also included Load parametric modulators for the six Speech × Task conditions, based on the dots on the screen. For LISTEN trials, the parametric modulators only captured the number of dots on the screen (i.e., visual load), whereas for TRACK trials, these modulators also captured the effect of tracking load. These models also included run-specific modulators including the six spatial realignment parameters, as well as a run intercept and linear trend. Modulators were mean centered and not orthogonalized, allowing control modulators to compete for variance with task modulators. Because of the long TR (10 s; 9 s silent gap between successive scans) in our sparse acquisition design, we modeled trial activation using a finite-impulse response model without serial autocorrelations. Contrast maps for main effects and interactions were calculated at the subject level and tested against zero at the group level using a factorial partitioned-error repeated-measures ANOVA (Henson and Penny, 2003).

We analyzed participants' behavior using custom MATLAB (R2018a) scripts and the JASP program (0.8.3) for ANOVA and Bayesian analyses (using the default Cauchy prior). Note that Bayes factors (BF_{10}) < 0.33 provide moderate evidence supporting the null hypothesis (e.g., that two groups are the same; Jarosz and Wiley, 2014). Follow-up fMRI analyses were performed using MATLAB and JASP. For our follow-up interaction analyses, we used a second general linear model that included all 24 Speech × Task × Load conditions, along with our run-specific nuisance terms (see above). We followed-up omnibus ANOVAs with *post hoc t* tests, correcting for multiple comparisons with the Holm's procedure for sequential tests. Brain–behavior relationships were cross-validated by fitting a linear regression model to predict BOLD contrasts from behavior while holding out one participant at a time, using this model to predict each held-out participant's BOLD contrast from their behavior and then correlating the predicted and observed BOLD contrasts.

Data availability. All data and code are available on request.

Results

Task performance

During LISTEN trials, participants reported whether they understood the gist of each sentence (Fig. 1). Participants reported understanding almost all the intelligible speech trials (Clear, 98.1%; NV12, 97.9%; NV6, 93.8%; Fig. 2*A*) and almost none of the Rotated trials (5.3%). The proportion of sentences reported as gist understood across our three speech types was very similar to the proportion of words reported accurately by a separate group of pilot participants, offering convergent validity for our gist-based measure of intelligibility (Fig. 2*A*). Logistic mixed-effects regression revealed that gist scores differed among the three intelligible speech types ($F_{(1,21)} = 12.6$, $p = 0.002$). Whereas gist scores for NV12 and Clear were not significantly different ($t_{(21)} = 0.36$, $p = 0.72$), gist scores were higher for Clear than NV6 ($t_{(21)} = 4.10$, $p = 0.0005$, $d = 0.90$).

During TRACK trials, participants tracked 1, 3, 4, or 6 moving dots and then selected the member of the tracked set with a three-alternative forced choice (Fig. 1). Logistic mixed-effects regression reveal that as tracking load linearly increased, participants were less accurate ($t_{(21)} = -11.8$, $p = 9.5 \times 10^{-11}$, $d = -2.6$; Fig. 2*B*), from 94% accuracy for one dot to 60% accuracy for six dots (33% chance rate). Linear mixed-effects regression also revealed that as tracking load increased, participants' log reaction times increased on accurate trials ($t_{(21)} = 12.3$, $p = 6.8 \times 10^{-11}$, $d = 2.6$), likely reflecting decision difficulty.

After the scanning session, participants performed a surprise memory test, reporting whether sentences, presented one at a time on the screen, had been heard during the scanning session or whether they were new (Fig. 2*C*). Sensitivity (recognition performance) was above chance (d' = 0) in all conditions. Probit mixed-effects regression revealed that participants had better memory for sentences heard during LISTEN than during TRACK ($t_{(23)} = 4.77$, $p = 0.0001$, $d = 0.99$).

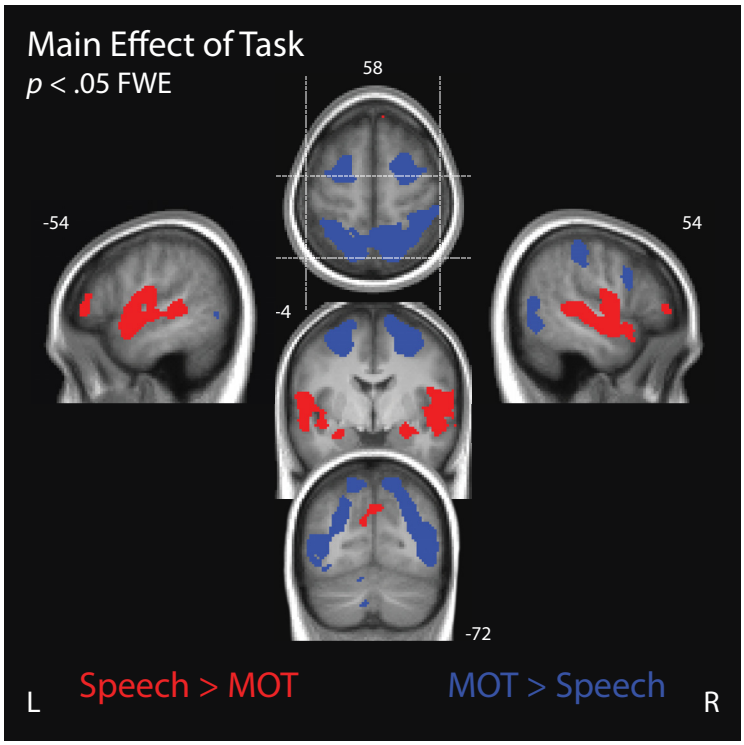

Figure 3. Main effect of Task. Voxels that exhibited a significant main effect of Task were colored according to whether they exhibited a greater response to LISTEN than TRACK or vice versa ($p < 0.05$, whole-brain FWE). Activation is plotted on the mean participant T1-weighted structural MR image, with dashed lines on the axial slice indicating the location of the sagittal and coronal slices. Extended Data Figure 3-1 shows the coordinates.

Recognition memory also significantly differed among Speech types ($F_{(2,21)} = 4.93$, $p = 0.018$), with marginally higher sensitivity for NV12 speech than Clear speech ($t_{(23)} = 2.07$, $p = 0.050$, $d = 0.43$). The interaction between Task and Speech type was only marginally significant ($F_{(2,46)} = 2.84$, $p = 0.10$). There was modest support for the specific interaction from previous research (Wild et al., 2012), replicating the finding that memory for NV6 speech suffered more from distraction than Clear speech ($t_{(23)} = 2.15$, $p = 0.043$, $d = 0.45$). These memory results suggest that performing the MOT task partially disrupted speech processing.

Task-specific neural responses

Participants appeared to orient their attention depending on the task cue (Fig. 3). Consistent with previous studies, LISTEN trials elicited greater activity across temporal and lateral prefrontal cortices (Scott et al., 2000; Davis and Johnsrude, 2003), whereas TRACK trials elicited greater activity in posterior parietal and superior frontal cortices (Culham et al., 2001; Howe et al., 2009).

We tested the simple main effect of Speech type during LISTEN trials only, as we hypothesized that speech processing would depend on attention (Fig. 4*A*). Comparing the activity elicited by Clear, NV12, NV6, and Rotated speech during LISTEN trials, we observed a simple main effect of Speech type across temporal and cingulo-opercular cortices. Temporal lobe voxels appeared to be sensitive to the intelligibility of speech, exhibiting progressively greater activity as gist report accuracy increased across the four speech types (green voxels; Davis and Johnsrude, 2003; Wild et al., 2012). In contrast, cingulo-opercular voxels exhibited greater activity for NV6 speech than for clear and NV12 speech (blue voxels), consistent with these regions responding more when stimuli are degraded (Wild et al., 2012; Eckert et al., 2016). These hypothesis-driven contrasts were not exhaustive, and some regions showed a main effect of speech with a different pattern of activation (white voxels).

Despite the highly similar intelligibility of Clear and NV12, our neural measures distinguished these speech types. Contrasting Clear versus NV12 during LISTEN revealed a significant peak in the left superior temporal gyrus (STG; $F_{(1,23)} = 80.46$, $p < 0.001$, whole-brain FWE) and a marginally significant peak in the right STG ($F_{(1,23)} = 40.91$, $p = 0.069$). These clusters partially overlapped with intelligibility-sensitive regions. Both STG regions were more sensitive to Clear than to NV12 speech. No voxels exhibited a significantly stronger response to NV12 than to Clear.

Finally, we tested for the simple parametric effect of tracking load during TRACK. In many of the regions that were more active for TRACK than LISTEN (main effect of task), BOLD activity was positively correlated with tracking load (Fig. 4*B*, green voxels), consistent with previous reports (Culham et al., 1998, 2001; Jovicich et al., 2001; Tomasi et al., 2004; Howe et al., 2009; Bettencourt, 2010). We also observed negative correlations with tracking load in the left supramarginal gyrus and angular gyri bilaterally (Fig. 4*B*, magenta voxels).

Domain-general response in anterior insulae

Our primary hypotheses concern the degree to which speech processing requires attention under different levels of degradation. Accordingly, we tested our two- and three-way interactions within a large speech-sensitive mask derived from previous data (Wild et al., 2012), which is fully independent of the current experiment (i.e., avoids using the same data for selection and analysis, so-called double dipping; Kriegeskorte et al., 2009). We defined our mask as voxels exhibiting either a significant main effect of Speech type or a Speech Type × Task interaction in this previous experiment (Wild et al., 2012, their Figs. 4, 5).

We observed a significant interaction between Task (LISTEN and TRACK) and Speech Type (Clear, NV12, and NV6) in the anterior insulae bilaterally, consistent with our previous experiment (Wild et al., 2012; Fig. 5). To compare the response profiles across hemispheres, we ran a Hemisphere × Speech Type × Task mixed ANOVA on the parameter estimates from these regions. The hemisphere factor did not influence our interaction effect ($BF_{10} = 0.201$), so we averaged parameter estimates across above-threshold voxels in this region across hemispheres.

In this insular region was a simple main effect of Speech type during LISTEN ($F_{(1.87,43.1)} = 18.65$, $p < 0.001$) that was not significant during TRACK ($F_{(1.53,35.2)} = 0.458$, $p = 0.585$; $BF_{10} = 0.172$; Fig. 5*A*). During LISTEN, the response of the anterior insulae was greater for NV6 than Clear speech ($t_{(23)} = 5.81$, $p_{Holm} < 0.001$, $d = 1.2$), and NV12 speech ($t_{(23)} = 5.10$, $p_{Holm} < 0.001$, $d = 1.1$). Activation during LISTEN for Clear and NV12 speech did not differ ($p_{Holm} = 0.229$; $BF_{10} = 0.423$). This pattern of elevated activity for difficult-to-understand degraded speech (NV6), only when this speech is task relevant, is consistent with the response profile observed in Wild et al. (2012).

To further characterize the task-dependent role of the anterior insulae, we also tested whether the effect of tracking load was evident in these insular voxels (Fig. 5*B*). We found that the insular response linearly increased with Load during TRACK

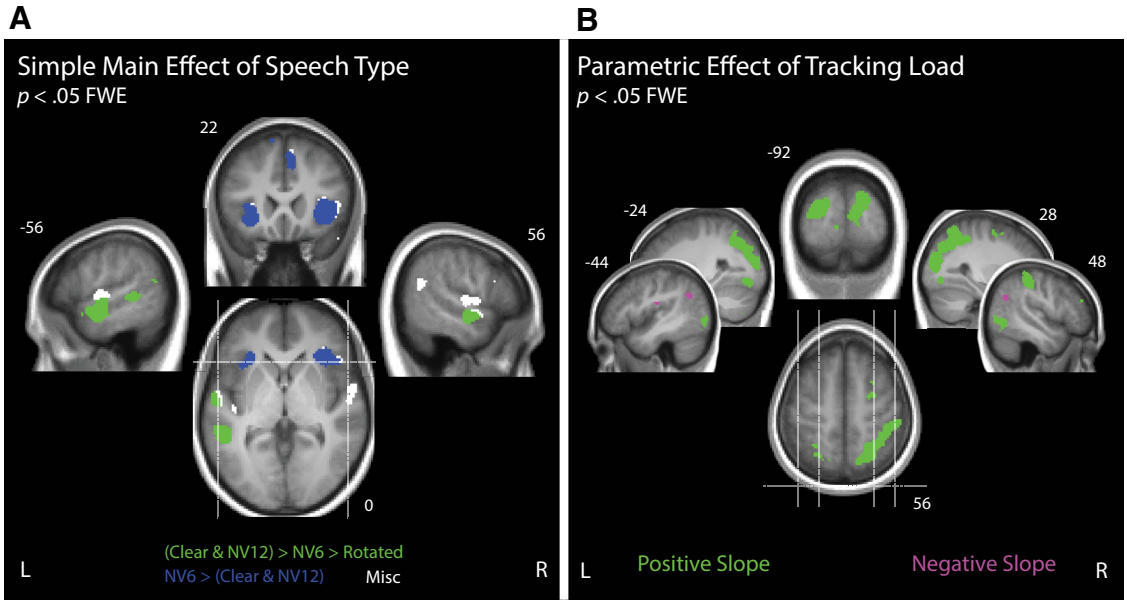

Figure 4. Task-specific simple main effects. ***A***, Simple main effect of speech type. Voxels that exhibited a significant simple main effect of Speech Type (Clear, NV12, NV6, or Rotated) during LISTEN are colored according to hypothesized contrasts (Wild et al., 2012). Green voxels indicate a greater response for more intelligible speech, and blue voxels indicate a greater response for NV6 compared with more intelligible Clear and NV12 speech. (White voxels exhibited any simple main effect pattern not captured by these contrasts.) ***B***, Parametric effect of tracking load. Voxels that exhibited a significant parametric effect of the number of dots tracked during TRACK are colored green if they show a positive relationship and magenta if they show a negative relationship. In both images, activation is shown on the mean participant T1-weighted structural MR image, and dashed lines on the axial slice indicate the location of the sagittal and coronal slices. Extended Data Figure 3-1 shows the coordinates.

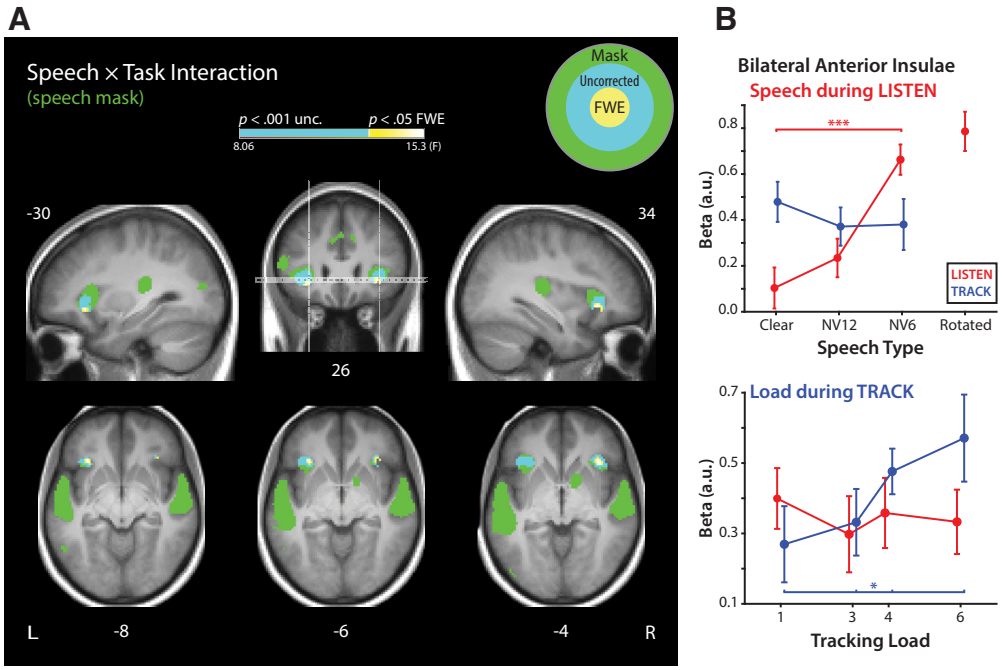

Figure 5. Speech × Task interaction. Analyses were performed within an independent mask of speech-sensitive cortex (green; see 'Domain-general response in anterior insulae'). Cyan voxels exhibited an interaction between Speech Type and Task at an uncorrected threshold ($p < 0.001$). Voxels that exhibited a significant interaction at a corrected threshold are indicated with a heat map corresponding to their F statistic ($p < 0.05$, within-mask FWE). ***A***, Parameter estimates extracted from above-threshold voxels show a significant simple main effect of Speech type only during LISTEN (red). ***B***, A *post hoc* analysis found a significant positive parametric effect of Load only during TRACK (blue). Error bars indicate SEM adjusted for within-subject measurements (Morey, 2008). Activation is plotted on the mean participant T1-weighted structural MR image, and dashed lines on the coronal slice indicate the location of the sagittal and axial slices. Extended Data Figure 3-1 shows the coordinates.

($t_{(23)} = 2.22$, $p = 0.036$, $d = 0.46$), with a stronger Load effect during TRACK than LISTEN ($t_{(23)} = 2.55$, $p = 0.018$, $d = 0.53$). Together, these signals suggest that the response of the insulae reflected the performance of the currently attended task.

Domain-specific response in anterior temporal cortex

Our analysis of primary interest examined whether there are speech-sensitive regions in which the effect of Speech type depends on the load during TRACK trials, and in particular

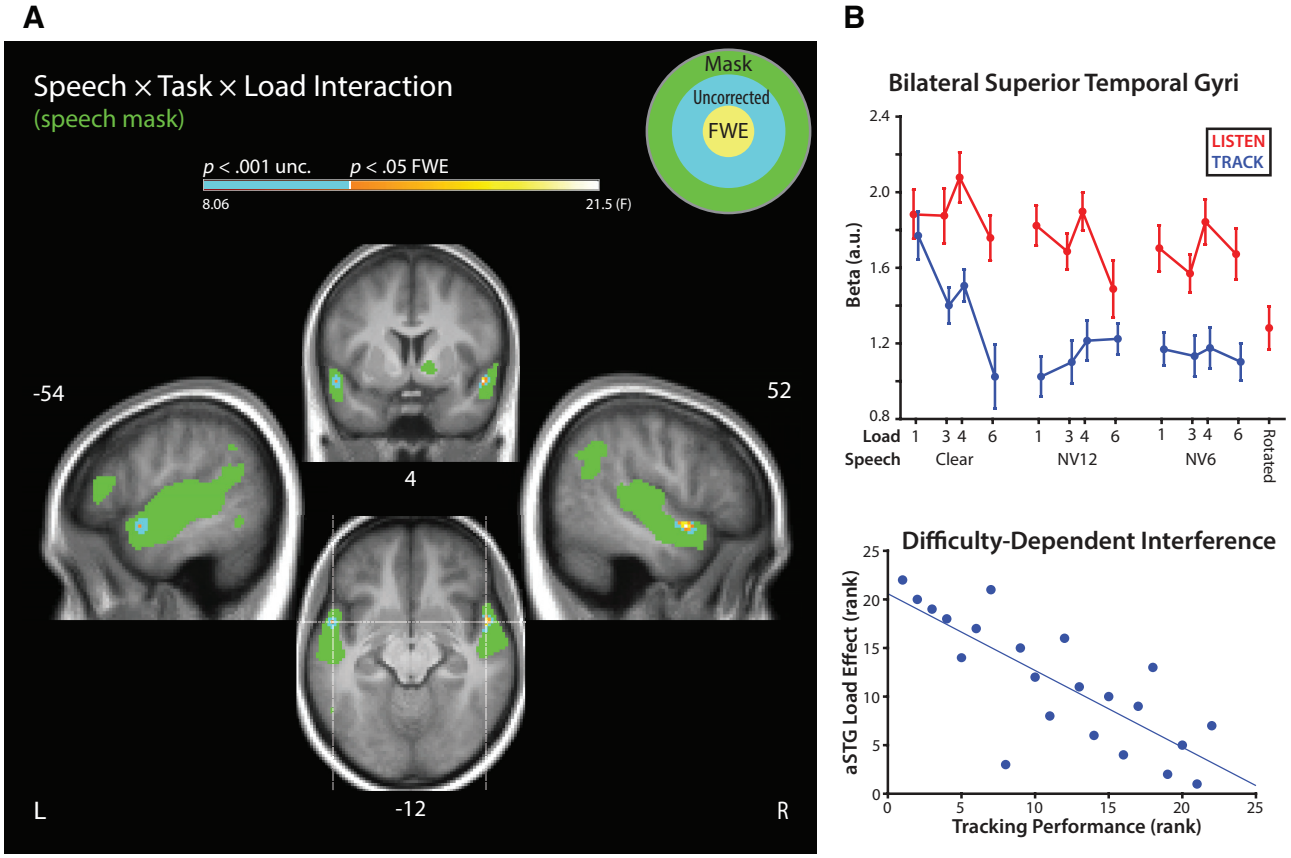

Figure 6. Speech × Task × Load interaction. Analyses were performed within an independent mask of speech-sensitive cortex (green; see 'Domain-general response in anterior insulae'). In cyan voxels, the slope relating BOLD activation to tracking load depended on both Task and Speech Type ($p < 0.001$, uncorrected). Voxels that exhibited a significant interaction at a corrected threshold are indicated with a heat map corresponding to their *F* statistic ($p < 0.05$, within-mask FWE). ***A***, Parameter estimates extracted from above-threshold voxels show a different load response for Clear and degraded speech during TRACK (blue), with degraded speech yielding activation in these regions at floor level (defined by the Rotated-speech point) at all tracking loads. In marked contrast, activity for Clear speech did not depend on task when Load was low (1-item MOT), but then linearly declined with increasing tracking load. ***B***, Participants who had more difficulty with the tracking task (lower Accuracy/RT) had a stronger interaction between Speech type and Load during TRACK. Error bars indicate within-subject SEM (Morey, 2008). Activation is plotted on the mean participant T1-weighted structural MR image, and dashed lines on the coronal slice indicate the location of the sagittal and axial slices. Extended Data Figure 3-1 shows the coordinates.

whether this cognitive load dissociates processing of Clear speech from intelligibility-matched degraded speech (NV12). Using the same speech-sensitive mask as our Speech Type × Task analysis, we examined the interaction of Speech Type × Task on the parametric Load modulators (effectively examining the Speech × Task × Load interaction). We found that this interaction was significant in anterior portions of the superior temporal gyrus (aSTGs; Fig. 6) bilaterally. As with the insulae, we found that this interaction was similar across hemispheres ($BF_{10} = 0.301$), so we averaged the parameter estimates across above-threshold voxels in both hemispheres.

During LISTEN, the effect of Load was not significant, nor was there a Load × Speech type interaction ($F_{(2, 46)} = 1.38$, $p = 0.267$, $BF_{10} = 0.278$). This was expected, since Load predictors during LISTEN only indexed the number of (task-irrelevant) dots on the screen. In contrast, during TRACK, the parametric Load effect depended on Speech Type ($F_{(2, 46)} = 12.13$, $p < 0.001$; Fig. 6*A*). The Load effect was apparent for Clear speech, with activity decreasing as load increased beyond 1-item MOT. In contrast, for NV12 and NV6 speech, activity during TRACK was at floor even for 1-item MOT, eliciting a response no stronger than for unintelligible rotated speech ($Load_{Clear} - Load_{NV12}$, $t_{(23)} = -4.04$, $p_{Holm} < 0.001$, $d = 0.84$; $Load_{Clear} - Load_{NV6}$, $t_{(23)} = -2.92$, $p_{Holm} = 0.016$, $d = 0.61$). Across all the Speech conditions in both tasks, only Clear speech during TRACK exhibited a significant effect of Load (Clear during TRACK, $t_{(23)} = -3.20$, $p_{bonferroni} = 0.024$, $d = 0.67$; all other $p_{uncorrected} \geq 0.16$ and $BF_{10} \leq 0.545$).

Another way to compare ourSpeech conditions is to examine, within each Speech Type, the MOT load at which differences between tasks begin to arise. Within each Speech Type, therefore, we compared the response during TRACK at each level of Load to that during LISTEN, averaged across levels of Load (as Load is irrelevant for this task). When one target was being tracked (lowest load), the STG response for clear speech was similar between TRACK and LISTEN ($t_{(23)} = -1.02$, $p_{uncorrected} = 0.32$, $BF_{10} = 0.344$; Fig. 6*A*, compare blue and red dots at Load level 1). In marked contrast, activity evoked by degraded speech depended strongly on Task; activity for both NV12 and NV6 was substantially lower during TRACK than LISTEN, even when only tracking one target ($NV12_{1\text{-target}}$, $t_{(23)} = -6.07$, $p_{Holm} < 0.001$, $d = -1.3$; $NV6_{1\text{-target}}$: $t_{(23)} = -5.76$, $p_{Holm} < 0.001$, $d = -1.2$). When tracking three or more objects, STG activity was always lower for TRACK than LISTEN and did not differ among speech types (Speech Type × Task when Load > 1, $BF_{10} = 0.038$). These results did not qualitatively change when comparing within specific load levels during LISTEN.

Complementing our neural measures, we also examined whether individual differences in the strength of this Load by

Speech type interaction was correlated with participants' task performance. We found that participants with a stronger aSTG Load effect during TRACK ($Load_{(NV12,\ NV6)} - Load_{Clear}$) had worse average overall tracking accuracy (Spearman's correlation, $\rho_{(20)} = -0.46$, $p = 0.032$) and slower median reaction times ($\rho_{(20)} = 0.52$, $p = 0.014$). We validated the generalizability of these individual differences using a leave-one-out cross-validation procedure. A measure of processing efficiency (accuracy/RT) was strongly correlated with aSTG Load effects within sample ($\rho_{(20)} = -0.79$, $p < 0.001$; Fig. 6*B*), and regression predictions for held-out participants strongly correlated with their performance ($\rho_{(20)} = 0.74$, $p < 0.001$). Participants with stronger neural indicators of load-dependent interference on speech processing performed more poorly on the MOT task, suggesting that our aSTG neural measures reflect the subjective task demands.

In sum, the response to clear speech in anterior temporal cortex was similar regardless of the focus of attention when tracking was easy but linearly declined to the same low level as for degraded speech with increasing tracking load. This neural index of interference was more severe for participants that were overall worse at the tracking task. The response profile for clear speech was fundamentally different from that for equally intelligible degraded speech, with activity for this degraded speech at the same level as unintelligible Rotated speech, even at weakest level of tracking load.

Discussion

Intelligibility responses in the anterior portion of the ventral speech pathways depend on attention (Sabri et al., 2008; Wild et al., 2012; Eckert et al., 2016). In the current experiment, we found that these regions can be fractionated based on whether speech sensitivity depends on the current task or the available processing capacity. Activity in the anterior insulae appeared to reflect the demands of the instructed task. This region responded more strongly to more degraded speech only when speech was task relevant, and activity depended linearly on tracking load only during MOT (Fig. 5). In contrast, sensitivity to speech in anterior temporal regions depended both on the type of speech and, for clear speech, on concurrent cognitive demands (Fig. 6). This load-dependent response in bilateral temporal lobes strongly dissociated clear speech from intelligibility-matched degraded speech. Clear speech was unaffected by the weakest level of distraction, at which the degraded speech response was already reduced to baseline. These observations functionally parcellate speech-sensitive cortex in the inferior frontal and superior temporal regions based on their relationship to cognitive control, demonstrating substantial costs of distraction under natural, perfectly intelligible, levels of speech degradation.

The anterior insulae play an important role in cognitive control (Duncan and Owen, 2000; Bunge et al., 2002; Dosenbach et al., 2006; Fedorenko et al., 2013; Shenhav et al., 2013; Cieslik et al., 2015), and may support performance monitoring (Wager et al., 2005; Vaden et al., 2013; Lamichhane et al., 2016), and/or orienting toward salient events (Klein et al., 2007; Seeley et al., 2007; Craig and Craig, 2009; Ullsperger et al., 2010). In this experiment, activity in the anterior insulae was sensitive only to the demands of the instructed task; stronger responses to degraded speech only during LISTEN (Wild et al., 2012) and positive linear dependence on tracking load only during TRACK. During LISTEN, this region exhibited a similar response for clear and intelligibility-matched degraded speech, also consistent with a generic role for performance monitoring (Vaden et al., 2013, 2015, 2016).

In anterior temporal cortex, we found that speech sensitivity depends on the cognitive demands of a distracting task. When Clear speech was task irrelevant, the aSTG response linearly declined as tracking load increased, with a stronger decline predicting poorer tracking performance. This decline may reflect a decreased availability of attention to enhance speech perception or active suppression of this region to reduce interference, with both accounts implying shared capacity for speech perception and MOT (Broadbent, 1958; Kahneman, 1973). MOT is a relatively simple task designed to isolate attentional processes that index object locations (Pylyshyn and Storm, 1988; Cavanagh and Alvarez, 2005; Scholl, 2009), with recent theoretical (Franconeri et al., 2010) and computational (Srivastava and Vul, 2016) models proposing that a critical function of MOT is protecting target indices from interference (i.e., from swapping a target with a distractor; Pylyshyn, 2004). During speech perception, there may be analogous competition among phonological, lexical, and semantic candidates (e.g., multiple potential interpretations of a sound or word), which is exacerbated by degradation (Miller et al., 1951; Marslen-Wilson, 1987; Thompson-Schill et al., 1997; Luce and Pisoni, 1998; Rodd et al., 2002; Novick et al., 2005; Spivey et al., 2005; Zhuang et al., 2011). During both tasks, attention could plausibly be allocated in response to heightened uncertainty and competition (e.g., toward regions of target-distractor proximity in MOT or proximal phonological candidates during speech), a core process in domain-general cognitive control (Berlyne, 1957; Posner and Snyder, 1975; Miller and Cohen, 2001). We speculate that these capacity-limited processes may allow effective use of context to constrain perception (Thompson-Schill et al., 1997), a promising area for future research.

When attention was on the MOT task, the anterior temporal response to (task irrelevant) intelligible degraded speech was eliminated, which contrasted markedly with the response during task-irrelevant clear speech. This profile may reflect maxed-out processing capacity or additional functions that are unavailable under distraction (e.g., functions that are goal dependent). That processing capacity was entirely occupied by the MOT task is not likely, given that the response in anterior temporal regions to mildly degraded speech was at the baseline even when individuals were tracking a single object, which is a very modest level of load. Furthermore, the load effect was clearly evident for task-irrelevant clear speech but not for degraded speech.

Instead, the processing of perfectly intelligible degraded speech in anterior temporal lobe regions appears to be gated by task goals. Consistent with this idea, activity in anterior insulae was determined by the demands of the attended task, plausibly in the service of top-down control over anterior temporal cortex (Novick et al., 2005; Wild et al., 2012; Eckert et al., 2016). The insulae and anterior temporal lobe share extensive anatomic connections via the uncinate fasciculus and extreme capsule (Petrides and Pandya, 1988, 2007; Romanski et al., 1999; Kier et al., 2004), which have long been thought to facilitate speech perception (Wernicke, 1908). Neuropsychological and neuroimaging evidence supports a role for this network in semantic processing (Hickok and Poeppel, 2007; Saur et al., 2008; Dick and Tremblay, 2012). For example, intracranial electrical stimulation of extreme capsule fibers in the anterior insulae reliably induces semantic paraphasias, with patients replacing target words with semantically related competitors (e.g., brush → comb; Duffau et al., 2005), a potential complement to the target-distractor swaps that characterize MOT

performance (Pylyshyn, 2004; Franconeri et al., 2010; Srivastava and Vul, 2016). Although these similarities are suggestive, our sparse-acquisition fMRI design limits our ability to test the role of frontotemporal connectivity on effortful speech perception. Further research using continuous fMRI, methods that target structural connectivity (e.g., diffusion-weighted imaging), or methods with higher spatiotemporal resolution (e.g., intracranial recordings) are needed to fully characterize the neural interactions that support selective attention during speech perception.

Consistent with enhanced top-down control during degraded speech perception, recognition memory tended to be better for NV12 speech than Clear speech when it was the focus of attention (Wild et al., 2012; Nairne, 1988; Hirshman and Mulligan, 1991). However, these findings are in contrast with previous research that has documented poorer memory for degraded speech (Rabbitt, 1966; Pichora-Fuller et al., 1995; Surprenant et al., 1999; Murphy et al., 2000). In many of these previous experiments, stimuli lacked the contextual constraints of full sentences (Rabbitt, 1966; Surprenant et al., 1999; Murphy et al., 2000), suggesting that the use of syntactic or semantic context to enhance speech intelligibility also enhances memory (Novick et al., 2005).

We found that task interference effects were strikingly different between clear and intelligibility-matched degraded speech, supporting an essential role for cognitive control at even the mildest levels of perceptual difficulty. These findings echo reports from individuals with hearing impairments that sustained perception of (amplified) speech is cognitive fatiguing. Nearly one in four people fitted with hearing aids report rarely using them, and one in five are neutral about, or dissatisfied with, their hearing aids (McCormack and Fortnum, 2013). The listening effort that is required to understand speech through hearing aids may be an important reason for this lack of enthusiasm. Our results demonstrate that even minor distractions during perception (i.e., tracking a single target) disrupts processing of mildly degraded speech; and this illustrates the need to consider cognitive load when assessing and accommodating listeners with hearing impairment.

References

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.

Berlyne DE (1957) Uncertainty and conflict: a point of contact between information-theory and behavior-theory concepts. Psychol Rev 64:329–339.

Bettencourt K (2010) Functional MRI and behavioral investigations of capacity limits in human visual attention. https://www.proquest.com/dissertations-theses/functional-mri-behavioral-investigations-capacity/docview/577653442/se-2?accountid=9758

Blesser B (1972) Speech perception under conditions of spectral transformation: I. Phonetic characteristics. J Speech Hear Res 15:5–41.

Broadbent DE (1958) Perception and communication. New York: Pergamon Press.

Bunge SA, Hazeltine E, Scanlon MD, Rosen AC, Gabrieli JDE (2002) Dissociable contributions of prefrontal and parietal cortices to response selection. Neuroimage 17:1562–1571.

Cavanagh P, Alvarez GA (2005) Tracking multiple targets with multifocal attention. Trends Cogn Sci 9:349–354.

Cieslik EC, Mueller VI, Eickhoff CR, Langner R, Eickhoff SB (2015) Three key regions for supervisory attentional control: evidence from neuroimaging meta-analyses. Neurosci Biobehav Rev 48:22–34.

Craig AD, Craig AD (2009) How do you feel–now? The anterior insula and human awareness. Nat Rev Neurosci 10:59–70.

Culham JC, Brandt SA, Cavanagh P, Kanwisher NG, Dale AM, Tootell RBH (1998) Cortical fMRI activation produced by attentive tracking of moving targets. J Neurophysiol 80:2657–2670.

Culham JC, Cavanagh P, Kanwisher NG (2001) Attention response functions: characterizing brain areas using fMRI activation during parametric variations of attentional load. Neuron 32:737–745.

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

Davis MH, Johnsrude IS, Hervais-Adelman A, Taylor K, McGettigan C (2005) Lexical information drives perceptual learning of distorted speech: evidence from the comprehension of noise-vocoded sentences. J Exp Psychol Gen 134:222–241.

DeCarlo LT (1998) Signal detection theory and generalized linear models. Psychol Methods 3: 186–205.

Dick AS, Tremblay P (2012) Beyond the arcuate fasciculus: consensus and controversy in the connectional anatomy of language. Brain 135:3529–3550.

Dosenbach NUF, Visscher KM, Palmer ED, Miezin FM, Wenger KK, Kang HC, Burgund ED, Grimes AL, Schlaggar BL, Petersen SE (2006) A core system for the implementation of task sets. Neuron 50:799–812.

Duffau H, Gatignol P, Mandonnet E, Peruzzi P, Tzourio-Mazoyer N, Capelle L (2005) New insights into the anatomo-functional connectivity of the semantic system: a study using cortico-subcortical electrostimulations. Brain 128:797–810.

Duncan J, Owen AM (2000) Common regions of the human frontal lobe recruited by diverse cognitive demands. Trends Neurosci 23:475–483.

Eckert MA, Teubner-Rhodes S, Vaden KI Jr (2016) Is listening in noise worth it? The neurobiology of speech recognition in challenging listening conditions. Ear Hear 37 Suppl 1:101S–110S.

Evans S, Davis MH (2015) Hierarchical organization of auditory and motor representations in speech perception: evidence from searchlight similarity analysis. Cereb Cortex 25:4772–4788.

Fedorenko E (2014) The role of domain-general cognitive control in language comprehension. Front Psychol 5:335.

Fedorenko E, Duncan J, Kanwisher N (2013) Broad domain generality in focal regions of frontal and parietal cortex. Proc Natl Acad Sci USA 110:16616–16621.

Franconeri SL, Jonathan SV, Scimeca JM (2010) Tracking multiple objects is limited only by object spacing, not by speed, time, or capacity. Psychol Sci 21:920–925.

Greenwood DD (1990) A cochlear frequency-position function for several species—29 years later. J Acoust Soc Am 87:2592–2605.

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.

Heald S, Nusbaum HC (2014) Speech perception as an active cognitive process. Front Syst Neurosci 8:35.

Henson RNA, Penny WD (2003) ANOVAs and SPM. Technical Report Wellcome Department of Imaging Neuroscience, London.

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

Hirshman E, Mulligan N (1991) Perceptual interference improves explicit memory but does not enhance data-driven processing. J Exp Psychol Learn Mem Cogn 17:507–513.

Howe PD, Horowitz TS, Morocz IA, Wolfe J, Livingstone MS (2009) Using fMRI to distinguish components of the multiple object tracking task. J Vis 9:10.

Jarosz AF, Wiley J (2014) What are the odds? A practical guide to computing and reporting Bayes factors. J Probl Solving 7:2.

Johnsrude IS, Rodd JM (2016) Factors that increase processing demands when listening to speech. In: Neurobiology of language (Hickok G, Small SL, eds), pp 491–502. Amsterdam: Academic Press.

Jovicich J, Peters RJ, Koch C, Braun J, Chang L, Ernst T (2001) Brain areas specific for attentional load in a motion-tracking task. J Cogn Neurosci 13:1048–1058.

Kahneman D (1973) Attention and effort. Englewood Cliffs, NJ: Prentice-Hall.

Kier EL, Staib LH, Davis LM, Bronen RA (2004) MR imaging of the temporal stem: anatomic dissection tractography of the uncinate fasciculus, inferior occipitofrontal fasciculus, and Meyer's loop of the optic radiation. AJNR Am J Neuroradiol 25:677–691.

Klein TA, Endrass T, Kathmann N, Neumann J, von Cramon DY, Ullsperger M (2007) Neural correlates of error awareness. Neuroimage 34:1774–1781.

Kriegeskorte N, Simmons WK, Bellgowan PSF, Baker CI (2009) Circular analysis in systems neuroscience: the dangers of double dipping. Nat Neurosci 12: 535–540.

Lamichhane B, Adhikari BM, Dhamala M (2016) The activity in the anterior insulae is modulated by perceptual decision-making difficulty. Neuroscience 327:79–94.

Luce PA, Pisoni DB (1998) Recognizing spoken words: the neighborhood activation model. Ear Hear 19:1–36.

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

McCormack A, Fortnum H (2013) Why do people fitted with hearing aids not wear them? Int J Audiol 52:360–368.

Miller EK, Cohen JD (2001) An integrative theory of prefrontal cortex function. Annu Rev Neurosci 24:167–202.

Miller GA, Heise GA, Lichten W (1951) The intelligibility of speech as a function of the context of the test materials. J Exp Psychol 41:329–335.

Morey RD (2008) Confidence intervals from normalized data: a correction to Cousineau. Tutorial in Quantitative Methods for Psychology 4:61–64.

Murphy DR, Craik FIM, Li KZH, Schneider BA (2000) Comparing the effects of aging and background noise on short-term memory performance. Psychol Aging 15:323–334.

Nairne JS (1988) The mnemonic value of perceptual identification. J Exp Psychol Learn Mem Cogn 14:248–255.

Novick JM, Trueswell JC, Thompson-Schill SL (2005) Cognitive control and parsing: reexamining the role of Broca's area in sentence comprehension. Cogn Affect Behav Neurosci 5:263–281.

Petrides M, Pandya DN (1988) Association fiber pathways to the frontal cortex from the superior temporal region in the rhesus monkey. J Comp Neurol 273:52–66.

Petrides M, Pandya DN (2007) Efferent association pathways from the rostral prefrontal cortex in the macaque monkey. J Neurosci 27:11573–11586.

Pichora-Fuller MK, Schneider BA, Daneman M (1995) How young and old adults listen to and remember speech in noise. J Acoust Soc Am 97:593–608.

Pichora-Fuller MK, Kramer SE, Eckert MA, Edwards B, Hornsby BWY, Humes LE, Lemke U, Lunner T, Matthen M, Mackersie CL, Naylor G, Phillips NA, Richter M, Rudner M, Sommers MS, Tremblay KL, Wingfield A (2016) Hearing impairment and cognitive energy: the framework for understanding effortful listening (FUEL). Ear Hear 37:5S– 27S.

Posner M, Snyder C (1975) Attention and cognitive control. In: Information processing and cognition: The Loyola symposium (Solso RL, ed), pp 55–85. Hillsdale, NJ: Erlbaum.

Pylyshyn Z (2004) Some puzzling findings in multiple object tracking: I. Tracking without keeping track of object identities. Vis Cogn 11:801–822.

Pylyshyn ZW, Storm RW (1988) Tracking multiple independent targets: evidence for a parallel tracking mechanism. Spatial Vis 3:179–197.

Rabbitt P (1966) Recognition: memory for words correctly heard in noise. Psychon Sci 6:383–384.

Rodd J, Gaskell G, Marslen-Wilson W (2002) Making sense of semantic ambiguity: semantic competition in lexical access. J Mem Lang 46:245–266.

Romanski LM, Bates JF, Goldman-Rakic PS (1999) Auditory belt and parabelt projections to the prefrontal cortex in the rhesus monkey. J Comp Neurol 403:141–157.

Rouault M, Koechlin E (2018) Prefrontal function and cognitive control: from action to language. Curr Opin Behavl Sci 21:106–111.

Sabri M, Binder JR, Desai R, Medler DA, Leitl MD, Liebenthal E (2008) Attentional and linguistic interactions in speech perception. Neuroimage 39:1444–1456.

Saur D, Kreher BW, Schnell S, Kümmerer D, Kellmeyer P, Vry M-S, Umarova R, Musso M, Glauche V, Abel S, Huber W, Rijntjes M, Hennig J, Weiller C (2008) Ventral and dorsal pathways for language. Proc Natl Acad Sci U S A 105:18035–18040.

Scholl BJ (2009) What have we learned about attention from multiple object tracking (and vice versa). In: Computation, cognition, and Pylyshyn (Dedrick D, Trick L, eds), pp 49–78. Cambridge, MA: MIT.

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.

Seeley WW, Menon V, Schatzberg AF, Keller J, Glover GH, Kenna H, Reiss AL, Greicius MD (2007) Dissociable intrinsic connectivity networks for salience processing and executive control. J Neurosci 27:2349–2356.

Shannon RV, Zeng F-G, Kamath V, Wygonski J, Ekelid M (1995) Speech recognition with primarily temporal cues. Science 270:303–304.

Shenhav A, Botvinick MM, Cohen JD (2013) The expected value of control: an integrative theory of anterior cingulate cortex function. Neuron 79:217–240.

Spivey MJ, Grosjean M, Knoblich G (2005) Continuous attraction toward phonological competitors. Proc Natl Acad Sci U S A 102:10393–10398.

Srivastava N, Vul E (2016) Attention modulates spatial precision in multiple-object tracking. Top Cogn Sci 8:335–348.

Surprenant AM, Neath I, LeCompte DC (1999) Irrelevant speech, phonological similarity, and presentation modality. Memory 7:405–420.

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.

Tomasi D, Ernst T, Caparelli EC, Chang L (2004) Practice-induced changes of brain function during visual attention: a parametric fMRI study at 4 Tesla. Neuroimage 23:1414–1421.

Ullsperger M, Harsay HA, Wessel JR, Ridderinkhof KR (2010) Conscious perception of errors and its relation to the anterior insula. Brain Struct Funct 214:629–643.

Vaden KI, Kuchinsky SE, Cute SL, Ahlstrom JB, Dubno JR, Eckert MA (2013) The cingulo-opercular network provides word-recognition benefit. J Neurosci 33:18979–18986.

Vaden KI, Kuchinsky SE, Ahlstrom JB, Dubno JR, Eckert MA (2015) Cortical activity predicts which older adults recognize speech in noise and when. J Neurosci 35:3929–3937.

Vaden KI, Jr, Kuchinsky SE, Ahlstrom JB, Teubner-Rhodes SE, Dubno JR, Eckert MA (2016) Cingulo-opercular function during word recognition in noise for older adults with hearing loss. Exp Aging Res 42:67–82.

Wager TD, Sylvester C-YC, Lacey SC, Nee DE, Franklin M, Jonides J (2005) Common and unique components of response inhibition revealed by fMRI. Neuroimage 27:323–340.

Wernicke C (1908) The symptom-complex of aphasia. Diseases of the nervous system (Church A, ed), pp. 265–324. New York: Appleton.

Wild CJ, Yusuf A, Wilson DE, Peelle JE, Davis MH, Johnsrude IS (2012) Effortful listening: the processing of degraded speech depends critically on attention. J Neurosci 32:14010–14021.

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.