7-29-2016 12:00 AM

The Effects of Concurrent Cognitive Load on the Processing of Clear and Degraded Speech

Harrison Ritz
The University of Western Ontario

Supervisor
Dr. Ingrid Johnsrude
The University of Western Ontario

Graduate Program in Psychology
A thesis submitted in partial fulfillment of the requirements for the degree in Master of Science

Recommended Citation
Ritz, Harrison, "The Effects of Concurrent Cognitive Load on the Processing of Clear and Degraded Speech" (2016). *Electronic Thesis and Dissertation Repository*. 4110.
https://ir.lib.uwo.ca/etd/4110

Abstract

A previous study has found that perceiving degraded speech requires attention, with compromised behavioral and neurological measures of speech processing for degraded speech, but not clear speech, when participants are distracted (Wild et al., 2012b). We extended these findings by examining behavioral and neural correlates of speech perception under different levels of cognitive load using multiple object tracking. We also investigated the role of attention in perceiving degraded speech that was as intelligible as clear speech, in order to separate perceptual outcomes (i.e., intelligibility) from the requisite processing demands. We found that the speech perception system is heterogeneous in its attentional requirements. The bilateral anterior insulae response reflected the cognitive load of the attended task, but not the unattended task, whereas activity in the anterior superior temporal gyrus reflected the cognitive load of both tasks. Under distraction, we found dissociable responses for clear and intelligibility-matched degraded speech.

Keywords: speech perception, degraded speech, attention, fMRI

Acknowledgements

I am deeply grateful the guidance that Ingrid Johnsrude has provided me over these past four years. She has always been generous in her mentorship, taking the time to teach an intellectual rigor that I will draw on again and again. Thank you so much for all the time and effort you have put into helping me develop as a careful, critical, scientist.

I am also grateful for the CoNCH lab for friendship and support. I would like to thank Julia, Chris, Bjorn, Emma, and Alex for providing me with guidance and support, and Rachel, Dora, Ysi, and Alenka for being close friends, although these roles are in no way mutually exclusive. I would also like to thank Natalie for her friendship, and for helping me collect the data for this experiment, as well as Jordan and Heather for supporting me during the final leg of this project.

Finally, I am grateful for my friends, my family, and Nishi, for their love and support.

Table of Contents

List of Figures

List of Appendices

The effects of concurrent cognitive load on the processing of clear and degraded speech

In perfect listening conditions, the comprehension of speech is seemingly effortless. However, everyday listening conditions are rarely as good as the laboratory, and speech is commonly degraded by noisy environments, by peripheral hearing impairment, or by low-fidelity digital communication. When speech is acoustically degraded, accurate perception places greater demands on cognitive resources than are required for the perception of clear speech (Heald & Nusbaum, 2014; Johnsrude & Rodd, 2016). Recent neuroimaging studies have found that attention modulates neural and behavioural responses to speech, and that focusing on speech is critical for comprehension in challenging listening conditions (Wild et al., 2012; Sabri et al., 2008). The current experiment will extend these observations by measuring neural indicators of the extent to which attention is required for the perception of clear and highly intelligible degraded speech.

Degraded speech perception

Degraded speech offers useful insight into the speech perception system, because it exaggerates the existing challenge of mapping highly variable acoustics on to stable linguistic categories (Liberman, Cooper, Shankweiler, Studdert-Kennedy, 1967; Heald & Nusbaum, 2014). The lack of invariance in the mapping between sounds and meaning (i.e., the bottom-up process) necessitates the use of context to help disambiguate speech (i.e., top-down biasing of perception), although there remains controversy over when and how context influences perception (for review, see: Heald & Nusbaum, 2014; Norris, McQueen, & Cutler, 2016). Acoustic degradation (such as noise-vocoding, which parametrically degrades the spectral clarity of speech, while largely preserving temporal information; Shannon et al., 1995) increases the uncertainty of linguistic mappings (in Bayesian terms, degradation broadens the likelihood distribution; Norris & McQueen, 2008), increasing the requirements for top-down control (i.e., a greater influence of priors; Norris & McQueen, 2008). In this respect, the *degradation* of a speech signal, i.e., the manipulation of its acoustic features, may be distinguished from the *distortion* of a speech signal, with the latter entailing the addition of spectral energy to a speech signal, such as masking with white noise or simultaneous speakers. This experiment primarily degraded speech (with minor distortion artifacts) in order to avoid the additional cognitive demands, such as stream segregation, that accompany a distortion manipulation. With our degradation manipulation, we instead directly disrupt participants' ability to map the acoustic signal on to linguistic categories. There are also methodological advantages to studying degraded speech perception. People are highly proficient at understanding speech, and so by degrading speech, and bringing accuracy down from near-perfect performance, researchers can make more sensitive measurements of the underlying perceptual processes.

A body of research over the past 50+ years suggests that the top-down processes required for accurate perception of degraded speech depend on domain-general working-memory and attentional resources. Early work by Patrick Rabbitt (1968) found that memory for clear speech was more disrupted by the subsequent perception of degraded speech than it was by clear speech, suggesting that the rehearsal of earlier items was a process that shared capacity with the perception of degraded speech. Luce, Feustel, and Pisoni (1983) similarly found that simultaneous cognitive load disrupted the perception of degraded speech more than clear speech. Several subsequent studies have found that participants exhibit poorer performance on working memory tasks that involve acoustically degraded, but intelligible, words (Pichora-Fuller, Schneider, & Daneman, 1995; Burkholder, Pisoni, & Svirsky, 2005; Wingfield, Tun, & McCoy, 2005; Francis & Nusbaum, 2009; Piquado, Cousins, Wingfield, & Miller, 2010; Obleser et al., 2012; Amichetti et al, 2013). Finally, studies examining individual differences in degraded speech perception have found correlations between participants' ability to understand degraded speech and measures of attention and working memory (Akeroyd, 2008; Humes, Kidd & Lentz, 2013; Besser et al., 2013).

Functional magnetic resonance imaging (fMRI) experiments of speech perception have found a reliable set of frontal and temporal regions in which participants' blood-oxygen-level dependent (BOLD) responses depend on the intelligibility of speech (Scott, Blank, Rosen, & Wise, 2000; Davis & Johnsrude, 2003; Obleser et al., 2007; Obleser & Kotz, 2010; Obleser, Eisner, & Kotz, 2011; Davis, Ford, Kherif, & Johnsrude, 2011; Wild, Davis, & Johnsrude, 2012a, Wild et al., 2012b, Evans et al., 2014). Regions near primary auditory cortex are sensitive to the acoustic features of speech (i.e., BOLD signal depends on the manipulations used to produce degradation), whereas regions anterior and posterior to primary auditory cortex are more sensitive to the intelligibility of speech, largely independent of acoustic features (i.e., are form-independent; Davis & Johnsrude, 2003; Okada et al., 2010; Evans et al., 2014). Several studies have found that in the inferior frontal gyrus, responses to speech are driven by both degradation (Giraud et al., 2004; Hervais-Adelman, Carlyon, Johnsrude & Davis, 2012; Wild et al., 2012b) and, critically, the interaction between degradation and contextual constraint (Obleser & Kotz, 2010; Davis, Ford, Kherif, & Johnsrude, 2011; Wild et al., 2012a), consistent with this region having a role in top-down modulation of lower-level speech regions. Consistent with such a modulatory role, primate anatomical research has found that frontal and anterior temporal regions have distinct connectivity and histological profile from primary sensory cortices and connections place them at a tertiary or quaternary level of auditory cortical processing (Kaas, Hackett, and Tramo, 1999), Strong, reciprocal connections link frontal regions and non-primary auditory cortices in the temporal lobe (Hackett et al., 1999; Romanski et al., 1999).

Dual-task Interactions

A powerful way to measure the cognitive demands of a task, such as perception of degraded speech, is to study the disruption produced by having participants perform a simultaneous task that has cognitive processes in common with the primary task (Kahneman, 1973). Several experimenters have found that less intelligible speech results in poorer performance on a secondary visual target-monitoring task (Downs, 1982; Feuerstein, 1992; Wild et al., 2012b; Pals, Sarampalis, & Baskent, 2013; for general review on dual-tasks and speech intelligibility, see: Gosselin & Gagné, 2010). These experimenters provide behavioral evidence that speech processes involved in compensating for stimulus degradation are sensitive to simultaneous cognitive demands.

However, dual-task designs provide a coarse characterization of specific cognitive processes, due to the limited outcome measures for an interference effect. Whereas interference may broadly disrupt speech perception, the resultant changes often occur across a single dimension of behavior. For example, performing a simultaneous task may impair participants' memory for speech, but it is difficult to determine where in the speech perception system this disruption occurred. In order to separate the effects of interference across the speech perception system, we can instead measure the simultaneous changes in neural activity throughout the brain while participants perceive clear and degraded speech under conditions of full attention or distraction.

Two recent fMRI experiments examined how perception of degraded and clear speech proceeds under full attention to speech compared to when attention is elsewhere. In an experiment by Sabri and colleagues (2008), participants listened to sentences, pseudowords, or unintelligible speech-like sounds ('rotated speech'; Blesser, 1972), while either attending to speech, or performing a simple visual short-term memory task. BOLD responses in many brain regions depended on the speech type, but only when participants focused on speech. In particular, the bilateral inferior frontal gyri and the left anterior temporal lobe differentiated between normal speech and rotated speech only during attend-speech, and the left middle frontal gyrus and left middle temporal gyrus differentiated between words and pseudowords only during attend-speech. This experiment is the first to use neuroimaging to study attentionally dependent speech processes, however, methodological concerns limit interpretation. Sabri and colleagues (2008) did not examine the effect of degraded acoustic quality on attentional demands, and the simultaneous scanner noise during their protocol may have imposed additional segregation demands during speech perception.

A subsequent study in our laboratory by Wild and colleagues (2012b) provided further evidence for attentionally dependent speech perception. In this experiment, participants heard meaningful sentences that were either clear or degraded but still highly intelligible (in one degraded condition, ~90% of the words in the sentences could be correctly reported). On each trial, participants either focused on speech or focused on concurrently presented visual or auditory stimuli and performed a simple target detection task on these. In a subsequent recognition memory test, memory for clear sentences was similar regardless of the locus of attention, but memory for degraded speech was worse when participants' attention was not on speech than when it was on speech. In regions involved in form-independent speech perception, the BOLD response to different speech types depended on the locus of attention. In the left inferior frontal gyrus (LIFG) and bilateral anterior insulae, BOLD responses were elevated for degraded speech, but only under full attention, providing further evidence for a role for the IFG in the effortful enhancement of degraded speech. In the anterior and posterior superior temporal gyrus/sulcus (STG/STS), the response profile mirrored participants' memory scores, with a decreased response under distraction for degraded, but not clear, speech.

This experiment demonstrated that attention is required for the comprehension of even quite highly intelligible degraded speech. However, questions still remain about how sensitive different regions in the speech perception system are to distraction. A core theoretical distinction in the dual-task literature is between interference from a primary task that completely obliterates performance of the secondary task ('processing bottleneck'), or that interferes with performance of the secondary task in a graded fashion ('capacity-sharing'; Kahneman, 1973; Pashler, 1984, 1994, 1998; Navon & Miller, 2002; Tombu & Jolicœur, 2003, 2005). It is possible that the interference Wild and colleagues (2012b) observed for degraded speech was due to a processing bottleneck in the speech perception system, such that when participants were distracted, these regions were simply not engaged in speech perception (c.f., Pashler, 1984, 1994, 1998). An alternative explanation is that speech processes share cognitive resources with the distractor task, and the allocation of cognitive resources to the visual task came at the expense of resources that were available for speech perception. This is proposed in 'resource pool' or 'effort' models of attention, which suggest that speech processing could access critical processes in parallel with the distractor task (c.f., Kahneman, 1973; Navon & Miller, 2002; Tombu & Jolicœur, 2003, 2005). While a bottleneck account is consistent with the interference observed in Wild et al. (2012b), in this experiment participants were not required to perform two tasks at the same time – they were required to attend to one task or to the other. However, the notionally unattended stimulus dimension may still be processed somewhat and that is what we set out to examine (c.f., Kahneman, 1973).

In order to characterize the how different brain regions involved in speech processing are affected by distraction, we will measure neural and behavioral responses to clear and degraded speech either during full attention, or during the performance of a distractor task that imposes a parametrically varied cognitive load. Speech regions that have a processing bottleneck will be sensitive to different speech types only when speech is the focus of attention, and will not differentiate between speech types under distraction, regardless of the difficulty of the secondary task. These regions may also be sensitive to the cognitive demands of the distractor task when it is the focus of attention. In such regions, speech processing is effectively gated by attention (i.e., these regions are processing bottlenecks). In speech regions that depend on a shared capacity, the response to different speech types should be differentially affected by simultaneous cognitive load. Such an interaction between the capacity demands for speech, and the capacity demands of the distractor task, would indicate that speech processing in such a region depends on shared cognitive resources. Both forms of interference may occur in different regions within the greater speech perception system, and neural responses to clear or degraded speech may be disrupted by either form of interference. However, we expect that the perception of clear speech will not be disrupted by distraction, at least when the load is modest: in Wild et al (2012b) manipulation of the locus of attention had little effect on neural and behavioral responses to clear speech, but the distractor tasks were not difficult (d′ scores on auditory and visual distractor tasks were 2.15 and 3.15 respectively; Wild et al., 2012b).

To further test the role of attention in speech perception, we will also examine the neural response to degraded speech matched to clear speech on intelligibility. This manipulation will test whether we can distinguish the outcomes of speech perception, i.e., intelligibility (measured, for example, as number of words from each sentence reported correctly), from the cognitive processes that are necessary to achieve this level of intelligibility. If capacity-sharing speech processes are involved in degraded speech perception, highly intelligible degraded speech should be more sensitive than less intelligible degraded speech to changes in cognitive load under distraction. This is because whatever capacity is remaining at low levels of tracking load may be insufficient for the perception of less-intelligible speech, whereas it may leave enough capacity to process highly intelligible speech. In addition, highly intelligible degraded speech allows us to determine which regions are sensitive to the outcomes of speech perception (i.e., depend on intelligibility), and which regions are sensitive to elevated processing demands (i.e., depend on stimulus degradation).

In order to measure whether there are speech processes that shared capacity under distraction, our secondary task must fit several criteria. An ideal task should have parametrically scalable attention demands, without requiring qualitatively different cognitive processes at different levels of load, in order to provide a parsimonious explanation for the manipulation of attentional demands. For example, in a 1-back condition of an n-back task (a popular working-memory paradigm; Kirchner, 1958) participants only need to remember the previous stimulus identity, whereas in a 2-back (or greater) condition participants must also inhibit the intermediate trials, qualitatively changing the nature of the task. An ideal task should also be non-verbal, to ensure that interference can be inferred to be arising from shared, domain-general, capacities (common to both speech listening and this nonverbal task). Finally, this task should also have a stable attentional requirement over the course of the speech stimulus, in order to minimize task-switching between modalities (c.f., Pashler, 1994). On the basis of these criteria, multiple object tracking is ideal for characterizing attentional processes in speech perception.

Multiple Object Tracking

Multiple Object Tracking (MOT) is a visual attention paradigm that provides some of the best current evidence for resource-pool models of attentional capacity. This task was originally developed to test whether a low-level form of visual attention, maintaining object indices (i.e., tracking objects) as they change location, operated in serial or parallel (Pylyshyn & Storm, 1988). In this task, a subset of identical dots are briefly highlighted as targets, the dots move pseudorandomly across a screen while participants track the targets, and then participants are tested on their knowledge of which dots were targets. Pylyshyn and Storm (1988) found that participants could simultaneously track several dots with a highly degree of accuracy, and that participants' performance was better than their best-case serial tracking model.

A prevailing theory of MOT is that it demonstrates attention as a flexibly allocated cognitive resource (for review see: Scholl, 2009; Cavanagh & Alvarez, 2005; Franconeri, Alvarez, & Cavanagh, 2013). Early models of MOT (including Pylyshyn & Storm, 1988) had a fixed number of indices that participants could assign to objects (typically 3-5; c.f. Cowen, 2001), framing MOT as the byproduct of a fixed set of tracking mechanisms (usually, 3-5) that operate in parallel, rather than a flexible cognitive resource that is sensitive to both the number of tracked objects, and the attentional demands of the task. Several subsequent studies have provided strong evidence for the resource-pool model of MOT (Alvarez & Franconeri, 2007; Bettencourt & Somers, 2009; Howe, Cohen, Pinto, Horowitz, 2010; Holcombe & Chen, 2012). In a particularly influential experiment, Alvarez and Franconeri (2007) found that the number of dots that participants could accurately track was highly dependent on the dots' velocity ($r^2 = .996$), and that participants could track up to eight items at slow velocities. The close relationship between tracking capacity and dot velocity is incompatible with slot models of attention, which hypothesize that attentional capacity is independent of the demands of the tracking task.

Neural responses during MOT provide further evidence that this task imposes parametric attentional demands. Experimenters reliably find that BOLD responses to MOT are linearly dependent on the number of targets that participants are tracking, particularly in the intraparietal sulcus, the superior parietal lobule, and the human frontal eye fields (Culham et al., 1998; Culham, Cavanagh, & Kanwisher, 2001; Jovicich et al., 2001; Tomasi, Ernst, Caparelli, & Chang, 2004; Howe et al., 2009; Tomasi, Wang, Wang, & Volkow, 2014). The strong modulation of BOLD responses by MOT suggests that, if speech perception shares cognitive resources with MOT, we should be able to detect load-dependent interference on speech perception.

Several behavioural experiments have demonstrated that MOT interferes with auditory attention tasks in dual-task designs. MOT appears to produce worse performance on a simultaneous auditory target detection task, relative to single-task performance (Alvarez et al., 2005). In another study, researchers observed that increasing the dot velocity during MOT interferes with an auditory target detection task (Tombu & Seiffert, 2006). Finally, Allen, McGeorge, Pearson, and Milne (2006) found that MOT performance was poorer when participants had to simultaneously categorize a tone as 'low' or 'high' compared to when MOT was performed alone. These results indicate that common processes are engaged during MOT and during performance of a difficult auditory perception task, suggesting that MOT may interfere with speech processing as well.

MOT meets our criteria for a task that may characterize the attentional demands of the speech perception system. Increasing the tracking load during MOT is thought to involve the flexible allocation of cognitive resources (Cavanagh & Alvarez, 2005), such that different levels of load involve qualitatively similar attentional processes. MOT is a spatial attention task that should not interfere with speech perception on the basis of stimulus similarity, or semantically categorizing trials as 'target present'. Finally, the attentional demands of MOT are consistent throughout the course of a trial, allowing for a constant load on speech perception over the course of a sentence.

Current Experiment

This experiment aims to measure how speech processes are affected by different attentional demands. We will measure participants' BOLD responses to clear and degraded sentences while they direct their attention towards either speech or MOT, allowing us to be sensitive to how processing of different qualities of speech, even when intelligibility is matched, changes as a function of attentional state. The structure of this experiment will be similar to that of Wild and colleagues (2012b). However, in this experiment, we will vary both the quality of the speech and the cognitive load of our distractor task. We will look for regions that differentiate between clear and even highly intelligible degraded speech in terms of attentional demands , as in Wild et al. (2012b). Specifically, we will examine whether, when attention is focused away from speech, processing of even highly intelligible degraded speech is completely obliterated in speech sensitive cortex, whereas clear speech is still processed. Uniquely in this experiment, we will also measure whether more demanding concurrent tasks can impair processing of even clear speech, which was largely unaffected by the distractor tasks used by Wild et al. (2012b). Finally, by comparing BOLD responses to stimuli that are matched in intelligibility, but differ in acoustic degradation, we may highlight processing dissociations related to the processes used to achieve high intelligibility. We predict that highly intelligible degraded speech, although matched to clear speech on intelligibility, will make greater attentional demands, reflecting recruitment of knowledge-guided processes required for enhanced intelligibility, compared to clear speech.

Predictions

We can make several specific predictions for our results, depending on how performance of a concurrent MOT task interferes with incidental processing of clear and degraded speech. For example, we would expect to replicate the findings of Wild et al. (2012b), who found an elevated response to degraded speech relative to clear speech in the anterior insulae when listeners attended to speech, and no differences among speech types when performing distractor tasks (Figure 1A). We might find that NV12 speech elicits a similar response as Clear (left), or NV6 (right), depending on whether the attention effects are late or early in speech perception, respectively.

We will also look for regions where the response depends on the combination of speech processing demands and tracking demands. We might find that the BOLD response to degraded speech depends on tracking load (i.e., BOLD activity during this condition negatively correlates with tracking load), but clear speech does not (no correlation between BOLD activity and tracking load; Figure 1B), with NV12 exhibiting a similar response to Clear (left) or NV6 (right). Finally, we might find that the BOLD response to clear speech depends on tracking load, but degraded speech does not (Figure 1C), again with NV12 exhibiting either a more similar response to Clear (left) or NV6 (right).

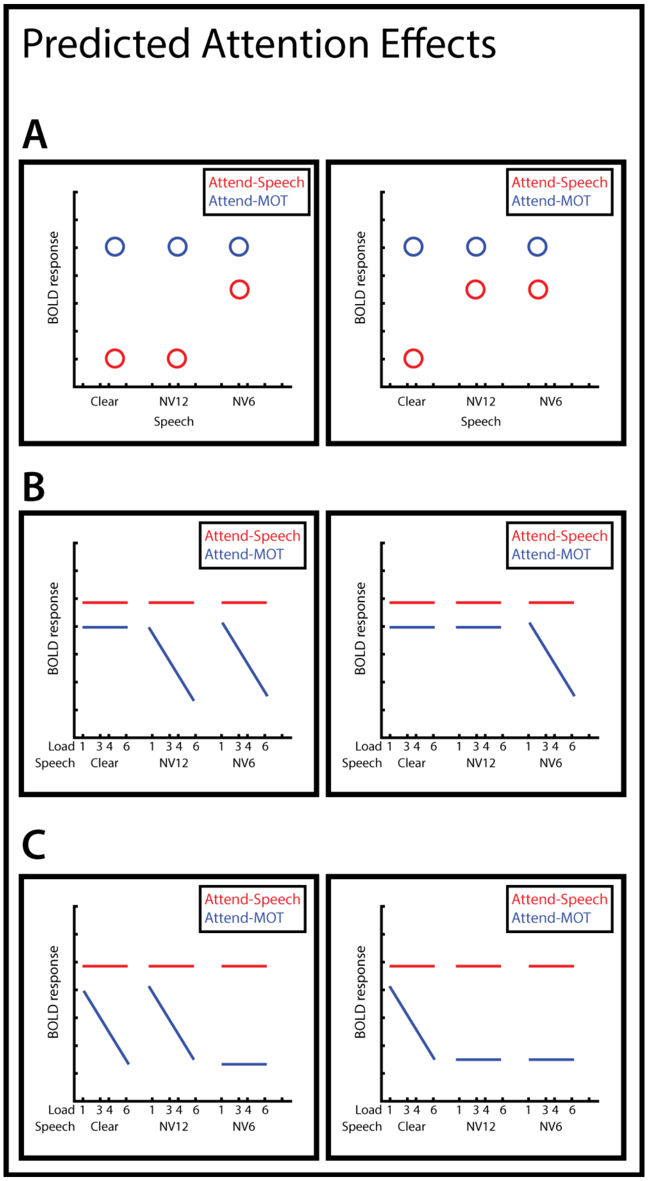

Figure 1. Predicted Attention Effects. We predicted that in different regions, activity evoked by different speech types would depend differentially on attention. One of the main research questions concerns the processing of nearly perfectly intelligible, but degraded sentences – are these processed like clear sentences are, or like more severely degraded sentences? In all rows, the left graphs depict a predicted response to NV 12 speech that is more similar to Clear (relative to NV6 speech), and the right graph depict a predicted NV12 response that is more similar to NV6 speech (relative to Clear speech). In A, circles represent the predicted mean BOLD response for each condition (averaged across the levels of Load), and in C and D, lines represent the predicted linear relationship between the BOLD response and Load within the other conditions. Note that, in the red Attend-Speech condition, the 'Load' factor simply indexes a minor variation in number of dots on the screen (between 13 and 16), and this is not predicted to influence activity (all red lines horizontal).

Methods

Participants

Twenty-six individuals (15 female; $M_{age} = 21.5$, $SD_{age} = 3.86$) participated in the fMRI portion of our experiment. Participants were right-handed, native English speakers (monolingually spoke English before the age of 5), had normal (or corrected-to-normal) vision, self-reported normal hearing, had no neurological or psychological disorders, and did not report losses of consciousness lasting longer than one hour. Participants were also screened on the basis of the MRI eligibility criteria at the Robarts Research Institute. We recruited participants via word-of-mouth and posters distributed on the campus of the University of Western Ontario. Two participants were removed before analysis, one due to technical issues with the stimulus delivery program, and another due to substantial movement during scanning (> 8mm translation in the z plane), leaving 24 for the analysis.

Twenty-four different individuals (19 female; $M_{age} = 21.1$, $SD_{age} = 2.11$) participated in piloting sessions for the behavioural portion of this experiment. Inclusion criteria and recruitment were the same as for the fMRI experiment. All participants provided full written consent, were debriefed following their session, and received monetary compensation for their participation. All experiments were cleared by the Health Science Research Ethics Board of the University of Western Ontario.

Experimental Design

In this experiment, we manipulated the task that participants performed, the clarity of speech that participants heard, and the number of dots that participants saw on a screen, in a fully factorial design. On each trial, participants both heard a sentence and saw moving dots. At the beginning of each trial, a word was presented on the screen, instructing them either to try to understand the speech (‘LISTEN’), or to perform a MOT task on the dots (‘TRACK’; see Figure 1). With three levels of speech clarity, and four levels of load, this experiment had 2 (Task) x 3 (Speech Type) x 4 (Load) conditions. Over 216 trials, participants experienced each condition 9 times, equally in each of the three runs. Since the load was manipulated parametrically, this design can also be viewed as a 6 condition (2 tasks x 3 speech clarity levels) experiment, with 36 trials in each condition over a range of dot densities. Baseline BOLD activity was measured over 24 silent, fixation-only trials. We also measured BOLD activity over 24 Attend-Speech trials with rotated NV speech, an acoustically matched unintelligible speech type that helps to localize speech-sensitive regions. Participants were instructed to respond to rotated trials

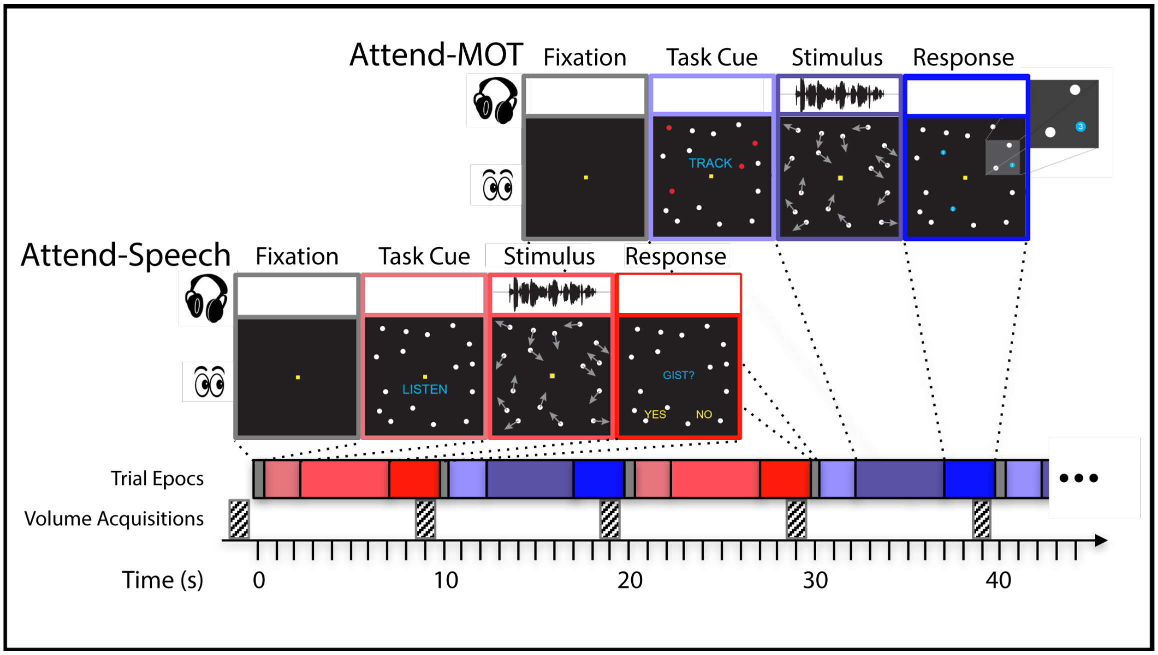

Figure 2. Trial Timecourse. On each trial, participants were first cued to either perform a tracking task (top, in blue), or attend to speech (bottom, in red), they then both heard speech and saw moving dots, and finally, a whole-brain acquisition was collected during their response. Timecourse: Participants first saw a fixation cross that indicated the trial onset (ITI: 300ms). Participants were then cued to either track the set of red-highlighted dots (tracking 1, 3, 4, or 6 dots) or focus on the speech (cue: 1.8sec). Next, participants both saw moving dots and heard an ordinary sentence (e.g., ‘Her handwriting was very difficult to read’), that was either clear (undistorted), 12-band noise-vocoded, or 6-band noise-vocoded (stimulus: 5sec). Finally, if participants attended to MOT, they made a three-alternative button-press indicating which one of three numbered dots they believed had been in the tracked set. If participants attended to speech, they reported, with a binary keypress, whether they had understood the gist of the sentence (response: 2.9sec). During this response, a whole-brain volume (TA = 1 sec) was acquired, with the onset of the scan occurring 4 seconds after the midpoint of the sentence.

as though they were regular Attend-Speech trials. These two types of control trial were randomly interspersed within runs, and occurred with equal frequency across runs.

Prior to the main experiment, participants completed two sets of training. The first set familiarized participants with noise-vocoded (NV) speech, in order for participants’ comprehension of NV speech to approximately reach asymptote. Over 24 trials, participants heard a sentence that was presented in NV12 or NV6 form, responded whether they had understood the gist of the sentence, and then received feedback by hearing the vocoded sentence again while the sentence was also written on the screen (as in Davis et al., 2005, Experiment 3). None of the sentences used during training were heard in the main experiment. During MOT training, participants practiced the tracking task for 24 trials. For the first half of training, participants performed a staircase version of the task, with the number of dots to be tracked increasing with each correct response, and decreasing with each incorrect response (within the range encountered during the experiment, i.e., tracking between 1 and 6 targets). In the second half of this training session, the tracking load changed randomly on each trial, with the constraint that two consecutive trials could not have the same tracking load.

After the 216 trials of the main experiment were completed, participants performed a recognition memory test for the sentences that they had heard. On each trial, participants saw a written sentence on the computer screen, and indicated with a keypress whether they remembered this sentence from the experiment (‘OLD’), or if it was a new sentence (‘NEW’). Participants made memory judgements on all 216 sentences from the task, along with 108 foil sentences. Foils sentences had slightly more words on average than target sentences (Foils: M = 10.3, SD = 2.15; Targets: M = 9.0, SD = 2.2), and differed from target sentences in both their topic and in all the content words they contained. Prior to the memory test, participants were unaware that memory for sentences would be tested, providing us with a measure of participants’ incidental encoding of these sentences when they were first heard.

Speech Stimuli

Over the course of the experiment, participants heard 216 everyday sentences (e.g., ‘His handwriting was very difficult to read.’; see Appendix A), all recorded from the same female speaker of Canadian English. Stimuli were presented diotically via foam-tipped insert earphones (Sensimetrics, Belmont, USA) at a comfortable listening level. The sentences were 6-13 words long; were 1.2 - 4.7 seconds in duration; and were split into six lists that were closely matched on the number of words (M = 9.0, SD = 2.2), the sentence duration (M = 2.5 sec, SD = 0.6 sec), and the logarithm of the summed word frequency (M = 5.5, SD = 0.2; Wilson, 1988). Each list was assigned to one of the six (two task by three speech types) conditions, counterbalanced across participants such that, across participants, each sentence was heard in each condition the same number of times.

The clarity of the speech stimuli was manipulated by noise-vocoding recorded sentences (Shannon et al., 1995). In this technique, a speech signal is partitioned into logarithmically spaced frequency bands, with boundaries chosen to be equally spaced along the basilar membrane (Greenwood, 1990). The amplitude envelope within each band is extracted (fourth-order Butterworth band-pass filter) and convolved with white band-limited noise sharing the same duration and frequency range. By changing the number of bands used in the process, we can make speech more or less intelligible (Shannon et al., 1995). In this experiment we used highly intelligible 12- and 6-band (NV12 and NV6) noise-vocoded speech. Piloting and previous experiments have found that NV12 and clear speech are closely matched at near perfect intelligibility, whereas the intelligibility of NV6 speech is poorer, but over 90% (see Figure 2). We generated spectrally rotated NV stimuli by reversing the order of the envelopes across frequencies, such that the envelope from the highest frequency range was applied to the lowest frequency range (and vice versa), and the envelope from the second highest frequency range was applied to the second lowest frequency range, etc (Blesser, 1972). The rotated speech samples that participants encountered in the experiment were generated using the sentences that participants heard during training. Since rotated speech is unintelligible, there was no expectation that participants would recognize these sentences.

Multiple-Object Tracking Task

Participants performed a multiple-object tracking task (MOT; Pylyshyn & Storm, 1988) at four different levels of tracking load. Participants tracked a subset of pseudorandomly moving dots, with tracking load manipulated by varying the number of dots that were tracked (and the total number of dots). At the beginning of each MOT trial, either 1, 3, 4, or 6 target dots amongst 12 distractor dots were highlighted for tracking by their colour changing from white to red for 1.8 seconds. All dots had a diameter of approximately 1 degree of visual angle, and were shown against a black screen spanning 20 x 20 degrees. After the cue, all dots started moving pseudorandomly around the screen at an approximate speed of 1.8 deg/sec, with dots repelling in the opposite direction from other dots or the edge of the screen at a 0.5-degree proximity. Participants were instructed to keep their gaze fixed on a static cue in the centre of the screen, and track the dots covertly, rather than with eye movements. After 5 seconds of tracking, dots froze in place, and three dots (one that had been tracked, and two foils) were highlighted with a blue colour and a number label (‘1’, ‘2’, or ‘3’). Participants had 2.8 seconds to indicate with a 3-alternative keypress which of the numbered dots was the one they had tracked, without feedback on their performance. There were 27 trials at each level of tracking load.

Pilot Experiments

In previous work, Conor Wild (Wild, 2012; Unpublished thesis, Section 2.2) used a subset of the stimuli used in the current experiment, and the same custom vocoder software. Fifteen young, normally hearing participants heard sentences presented one at a time and were asked to write down all the words they could understand from each sentence. Wild observed that the proportion of words reported correctly was similar for Clear sentences (M = .9802, 95% CI = [.9619 .9985]) and NV12 sentences(M = .9872, 95% CI = [.9721 1.0]), and lower for NV6 sentences (M = .9463, 95% CI = [.9248 .9679]; see Figure 2, white circles).

A set of pilot experiments was intended to confirm these intelligibility values, and to ensure that we were using levels of the MOT task that would result in off-ceiling performance that also reflected changes in tracking load. In these pilots, the procedure was similar to the fMRI experiment with both MOT and speech stimuli concurrently present, and participants performed either the MOT task (50% of the time) or attended to the speech. In a single-walled soundproof booth, sentence stimuli were delivered at a comfortable listening level via headphones (Grado Labs, Brooklyn, USA). We tested a few participants on many different versions of the MOT task, and so we will only report the intelligibility results from the 24 participants that took part in this pilot. As in the fMRI experiment, intelligibility was measured as the proportion of sentences that participants reported comprehending (see Figure 2, grey circles). We found that intelligibility was high and similar for Clear sentences(M = .9977, 95% CI = [.9879, 1.0]) and NV12 sentences (M = .9882, 95% CI = [.9762, 1.0]), and lower for NV12 sentences (M = .9304, 95% CI = [.9100, .9508]). These results accord well with the finer-grained word-report results of the first pilot, and provide further evidence, with a larger sample size, that the intelligibility of all stimuli was high, that the intelligibility was similar for Clear and NV12 speech, and that Clear and NV12 speech are more intelligible than NV6 speech. These results also suggest that our measures of intelligibility converge across finer (word report; pilot 1) and coarser (comprehension report; pilot 2) measures of intelligibility.

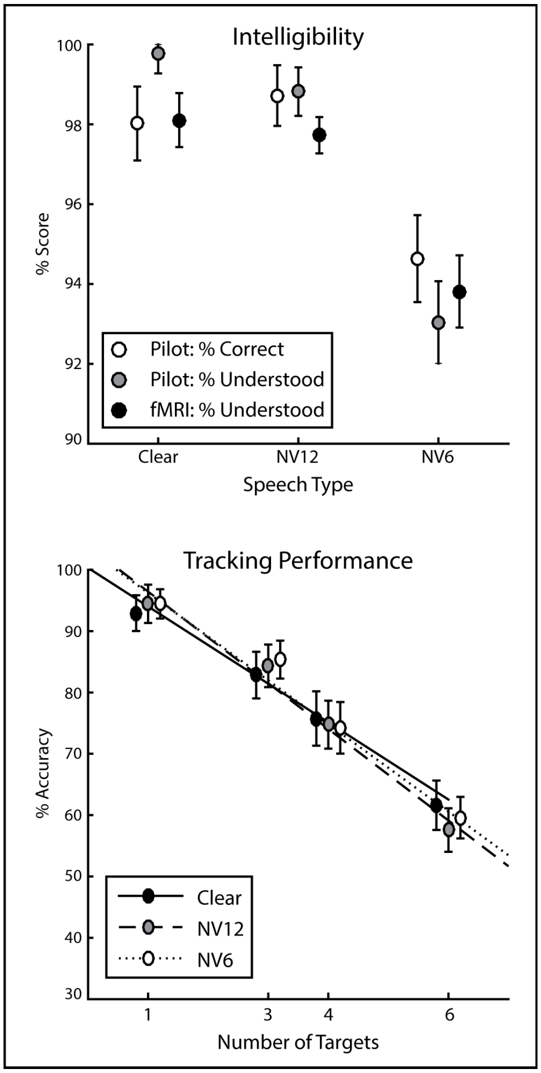

Figure 3. *In-Scanner Behavioural data.* Above: Intelligibility indices across pilot and fMRI experiments. During the first pilot experiment, we measured the intelligibility of our stimuli by measuring the proportion of words that participants could accurately report at each level of degradation (n = 12; white circles). With a separate set of participants, we measured the proportion of sentences for which participants reported comprehending the gist, while performing the same task as the fMRI participants (n = 24; grey circles). During the fMRI experiment, we again measured the proportion of sentences of each speech type for which participants reported comprehending the gist (n = 22; black circles). The subjective gist report scores closely matched the objective word report accuracies. Below: In-scanner tracking performance. As participants tracked more targets, their accuracy at indicating which of three dots had been a target declined. Participants performed above chance (33%) at all levels of tracking load, and their tracking performance was highly correlated with the number of targets (r_{median} = -.94). This correlation did not depend on the simultaneously heard Speech Type. For both graphs, errors bars indicate SEM adjusted for within-subject measurements (Morey, 2008).

fMRI Acquisition

Participants completed the fMRI experiment in a 3.0T Siemens Prisma MRI system at the Robarts Research Institute at the University of Western Ontario (London, Ontario, Canada). T2*-weighted functional images were acquired using an interleaved 4-factor multi-band EPI (field of view: 192mm x 192mm; resolution: 2.5mm isotropic; slice thickness: 2.5mm with a 10% gap; TE: 30ms; TA: 1sec per volume; TR: 10sec; flip angle: 70°). Transverse slices were acquired in order to obtain a whole-brain volume. To aid in spatial localization, T1-weighted structural images were collected at the beginning of each session using a single-shot EPI (field of view: 256mm x 256mm; resolution: 1mm isotropic; slice thickness: 1mm with a 50% gap; TE: 2.98ms; TR: 2300ms; flip angle: 9°).

Volumes were collected using a sparse acquisition protocol (Hall et al., 1999), in which speech stimuli were presented during the silent period (9 seconds) between scans. This protocol prevented scanner noise from stimulating auditory regions and further degrading our speech stimuli. Scans were acquired beginning 4 seconds after the mid-point of each sentence in order to sample the haemodynamic response close to the peak amplitude.

fMRI Preprocessing

fMRI Data were preprocessed and analyzed using Statistical Parametric Mapping (SPM12; Wellcome Centre for Neuroimaging, London, UK). We rigidly realigned functional images to the mean image of each run, and then coregistered participants' structural images to the mean functional image across runs. Next, we calculated an affine transformation and non-linear deformation (D'Agostino, Maes, Vandermeulen, & Suetens, 2004) for each structural image in order to match SPM12's default tissue probability maps (Fonov et al., 2009), segmenting and normalizing our images into MNI space. We applied these transformations to all of our functional images, and resampled them to a 2mm isotropic resolution. Finally, all volumes were spatially smoothed using a 3D Gaussian kernel with an 8mm FWHM.

fMRI Model

We constructed statistical parametric maps for each subject using a general linear model (GLM). Scans were modelled as occurring in one of eight trial types, corresponding to the six combinations of speech (3 levels) and task (2 levels), rotated speech trials, and silent baseline trials. We also included six parametric modulators, one for each combination of speech and task, corresponding to the number of dots that participants saw on the screen. Each of the three runs was modelled separately with 14 task predictors, six realignment parameters (to account for movement during scanning, both in terms of translation and rotation), and a predictor to remove the mean signal in each run. Due to the long TR (10 seconds), we did not model serial auto-correlations. Contrast maps for each main effect and interaction were calculated for each subject, and subjected to a group analysis using a factorial partitioned-error repeated-measures ANOVA (Henson & Penny, 2003).

Results

Behavioural Results

Due to a technical error, behavioural responses during scanning (i.e., intelligibility reports and tracking accuracies) were lost for 2 participants, leaving 22 participants for these analyses. Imaging data for these participants were still analyzed.

Intelligibility

During the speech task, participants indicated whether or not they understood the gist of each sentence with a binary 'yes/no' keypress with their dominant hand (see Figure 1). There was a significant main effect of speech type (Clear, NV12, and NV6) on intelligibility (One-way repeated-measures ANOVA: $F(1.37, 27.8) = 11.44$, $p = .001$, partial eta-squared = .353, all ANOVAs Greenhouse-Geisser corrected; see Figure 2, black circles). Pairwise t-tests (Šidák-corrected in order to control the family-wise type I error rate; Šidák, 1967) did not reveal a significant intelligibility difference between Clear and NV12 ($t(21) = 0.65$, $p = .893$), whereas participants reported significantly greater intelligibility for Clear than NV6 ($t(21) = 3.44$, $p = .007$), and greater intelligibility for NV12 than NV6 ($t(21) = 3.81$, $p = .003$). Thus, Clear and NV12 were similarly highly intelligible, and although NV6 was also highly intelligible, it was less intelligible than Clear or NV12. This pattern of intelligibility across speech types was similar to the pattern observed in two pilot experiments using the same stimuli and noise-vocoder, both in terms of the proportion of words participants could accurately recite from each sentence (n = 12), as well as the proportion of sentences that participants reported understanding (n = 24; see Figure 2).

Tracking Performance

For trials on which participants performed the tracking task, they tracked either 1, 3, 4, or 6 target dots amongst 12 identical distractor dots, and subsequently chose which of 3 dots they believed had originally been a target (see Figure 1). Participants consistently performed above the 33% chance rate, even when tracking the maximum number of dots (one-sample t-test against 33% for 6 dots: $t(21) = 10.95$, $p < .001$; see Figure 2). Participants' tracking accuracies were linearly dependent on the number of dots they tracked (Pearson's $r_{\text{median}} = -.94$, $r_{\text{IQR}} = .082$; one-sample t-test against 0 on Fisher z-transformed correlation coefficients: $t(21) = -11.43$, $p < .001$). We did not observe a significant difference in the strength of this linear dependence across speech types (one-way repeated-measures ANOVA on z-transformed correlation coefficients: $F(1.55, 32.5) = 0.405$, $p = .618$, partial eta-squared = .019).

Recognition Post-Test

Following the main experiment, participants performed a surprise recognition test. Participants judged all 216 sentences from the main experiment, and 108 foils that were matched for length but that did not overlap in content words with the targets, on whether they were present during the experiment. Sentences were presented visually, one at a time. We calculated d' to index participants' recognition of sentences heard previously in each speech (Clear, NV12, NV6) and task (Attend-Speech, Attend-MOT) condition (see Figure 3). Sensitivity (d') was above chance for all conditions (Šidák-corrected one-sample t-tests against 0: all ps < .001). Participants exhibited an overall conservative bias (bias to answer 'NEW'), likely due to the greater number of target trials than foil trials, whereas participants may have expected there to be a 50/50 split ($c_{\text{mean}} = 0.37$, $c_{\text{SD}} = 0.45$; One-sample t-test against 0: $t(23) = 4.07$, $p = .001$).

In order to examine the effects of speech type and task at encoding on subsequent memory, we ran a 3 x 2 (Speech Type; Clear, NV12, NV6 by Task; Attend-Speech,

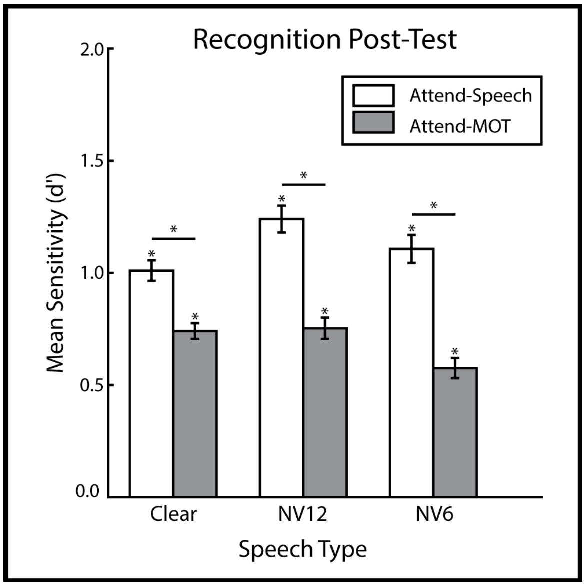

Figure 4. *Post-scan recognition memory performance*. Following the main experiment, participants decided whether sequentially presented written sentences had been in the experiment, or were novel. We used the proportion of sentences participants correctly reported as being from the experiment (hits) and the proportion of novel sentences that they misattributed as hearing during the experiment (false alarms) to calculate sensitivity (d') for each condition. Participants recognized sentences from all conditions better than chance (i.e., all d′s > 0; asterisk above error bar). Participants were better at remembering sentences they had heard during Attend-Speech than Attend-MOT in for every speech type (bracket with asterisk), with marginally greater differences between tasks for both NV speech types than for Clear speech. Errors bars indicate SEM adjusted for within-subject measurements (Morey, 2008).

Attend-MOT) repeated-measures ANOVA on participants' d' scores. We observed significant main effects for both the Speech Type ($F(1.91, 44.0) = 6.00$, $p = .006$, partial eta-squared = .207) and Task (Attend-Speech > Attend-MOT: $F(1, 23) = 85.84$, $p < .001$, partial eta-squared = .782). Although recognition of Clear and NV6 speech did not differ (all post-hoc contrasts Šidák-corrected paired t-test: $t(23) = .65$, $p = .891$), NV12 speech was recognized significantly better than both NV6 ($t(23) = 3.63$, $p = .004$) and Clear ($t(23) = 2.57$, $p = .048$).

We also found a significant Speech Type × Task interaction ($F(1.98, 45.5) = 3.97$, $p = .026$, = .15). Despite participants having better memory for all sentences heard during Attend-Speech than Attend-MOT for all speech types (all *p*s < .001), this difference was smaller for Clear than it was for both NV12 ($t(23) = -2.31$, $p = .091$) and NV6 ($t(23) = 2.54$, $p = .054$). Within each task, we found that during Attend-Speech, memory was better for NV12 than Clear speech ($t(23) = 2.82$, $p = .029$), whereas during Attend-MOT, memory was poorer for NV6 than both Clear speech ($t(23) = 2.79$, $p = .031$) and NV12 ($t(23) = 2.63$, $p = .044$, all contrasts Šidák-corrected). We did not observe correlations between tracking load and recognition memory for any speech type (one-sample t-test on Fisher's z-transformed correlation coefficients, all *p*s > .36).

fMRI Results

Main Effect of Task

We observed widespread activity that differentiated between the speech ('Attend-Speech') and tracking ('Attend-MOT') tasks (see Figure 4). Attend-Speech elicited greater activity across temporal and lateral prefrontal cortices, as predicted (Wild et al., 2012b), and Attend-MOT elicited greater activity in posterior parietal and superior frontal cortices, as predicted (Culham, Cavanagh, & Kanwisher, 2001; Howe et al., 2009). This activity pattern suggests that participants oriented their attention depending on the task cue.

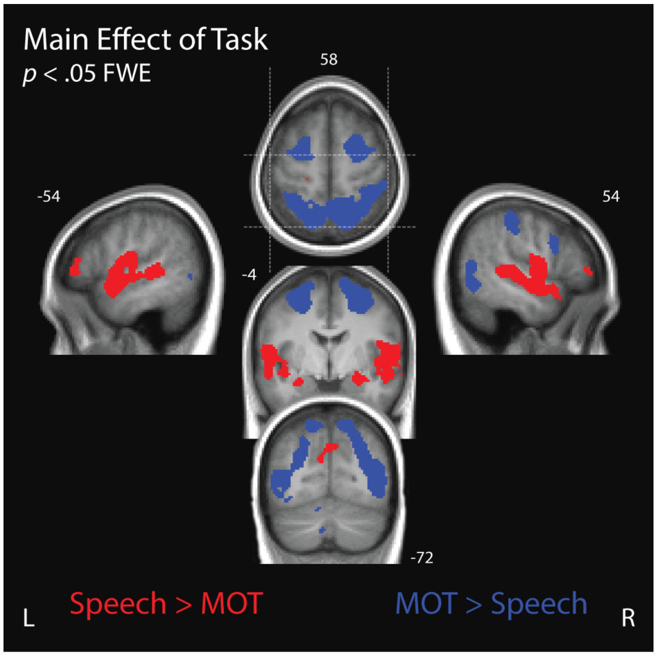

Figure 5. *Main effect of Task.* Voxels that exhibited a significant main effect of Task are grouped according to whether they reflect Attend-Speech BOLD activity > Attend-MOT (red) or Attend-MOT BOLD activity > Attend-Speech ($\alpha = .05$, corrected family-wise across the whole brain). Activation is plotted on the mean participant T1 image, and white dashed lines on the axial slice indicates the locations of the sagittal and coronal slices.

Simple Main Effect of Attended Speech Type

We analyzed our main effect of speech only for trials during which participants performed the speech task (i.e., simple main effect during Attend-Speech), since we hypothesized that focus of attention would alter speech processing; also, we wished to include rotated speech in the contrast as a baseline, but this never occurred during Attend-MOT (see Figure 5). Comparing the activity elicited by Clear, NV12, NV6, and Rotated speech during Attend-Speech, we observed a simple main effect of speech type across superior temporal and anterior insular cortices. Following Wild et al. (2012b), we examined intelligibility- and distortion-elevated simple effects within voxels exhibiting this simple main effect.

We first looked for voxels that had greater activity when speech was more intelligible (Figure 5, green voxels). To examine this, we first ranked our speech types by the mean gist comprehension reports across participants (combining Clear and NV12 due to their highly similar intelligibility). We then conducted a Helmert contrast (c.f., Wendorf, 2004), sequentially comparing the activity elicited by a higher level of intelligibility to the mean of all lower levels (i.e., [(Clear + NV12) > (NV6 + Rotated)] & [NV6 > Rotated]). Similar to Wild and colleagues (2012b), we observed intelligibility-elevated activity in bilateral superior temporal gyri.

Next we examined where activity was elevated for less intelligible (but still comprehensible) speech, compared to more completely intelligible speech, as such regions may be involved in compensating for stimulus degradation (Figure 5, blue voxels). Again, we combined our two high-intelligibility speech types (Clear and NV12), and searched for voxels where NV6 elicited greater activity than high-intelligibility speech (i.e., NV6 > (Clear + NV12)/2). As in Wild et al., (2012b), we found degradation-elevated activity in the anterior insulae and dorsal anterior cingulate cortex bilaterally.

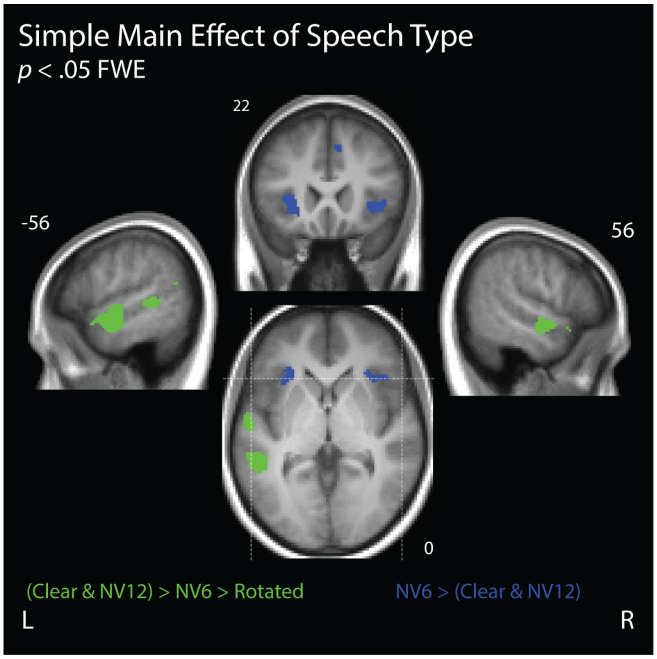

Figure 6. *Simple main effect of Speech Type.* Voxels that exhibited a significant simple main effect of Speech Type (Clear, NV12, NV6, or Rotated) during Attend-Speech are grouped by color: green indicates increasing BOLD activity with increasingly intelligible speech; and blue indicates greater activity for NV6 compared to more intelligible clear and NV12 speech (α = .05, corrected family-wise across the whole brain). Few voxels exhibited a simple main effect that was not captured by one of these two contrasts, and these are not shown. Activation is plotted on the mean participant T1 image, and white dashed lines on the axial slice indicates the locations of the sagittal and coronal slices.

Finally, we explicitly tested the prediction that, despite the highly similar intelligibility of Clear and NV12, we would observe dissociable neural responses to these speech types. The simple main effect of Clear vs NV12 during Attend-Speech revealed a significant peak in the left STG ($F(1, 23) = 86.49$, $p < .001$, FWE corrected across the whole brain) and a marginally significant peak in the right STG ($F(1, 23) = 42.68$, p = .054, FWE corrected across the whole brain). These clusters partially overlapped with voxels sensitive to intelligibility. In both STG regions, these effects were driven by a stronger response to Clear than NV12. No voxels exhibited a significantly stronger response for NV12 than for Clear.

Tracking-Load Dependent Activity

We looked for voxels in which activation was linearly dependent on the number of dots that participants were tracking (i.e., where BOLD signal correlated with MOT level; see Figure 6). In many of the brain regions in which there was greater activity for attend-MOT than attend-speech (main effect of task), there were also positive correlations between BOLD activation and the number of dots that participants tracked. These findings are consistent with previous studies examining the effect of MOT load using fMRI (Culham et al., 1998; Culham, Cavanagh, & Kanwisher, 2001; Jovicich et al., 2001; Tomasi, Ernst, Caparelli, & Chang, 2004; Howe et al., 2009). We also observed

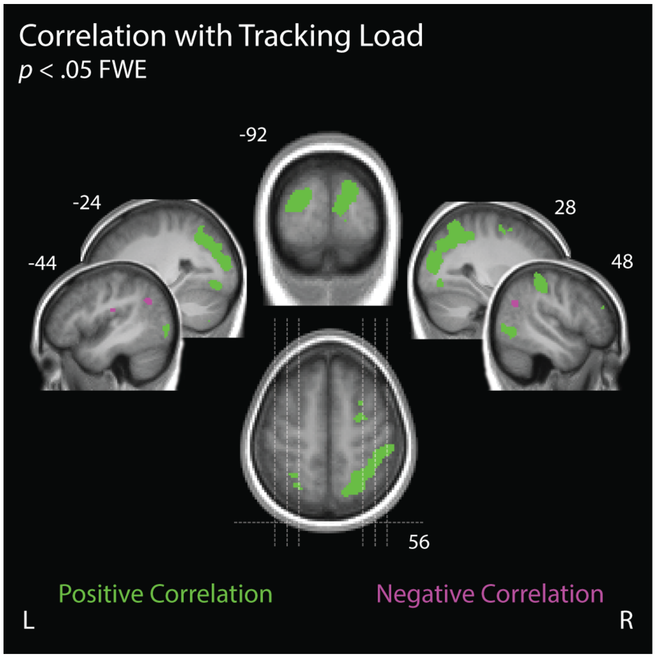

Figure 7. *Correlation with Tracking Load.* Voxels that exhibited a significant correlation with the number of dots that participants tracked during Attend-MOT are grouped by direction of effect: positive correlation is shown in green) and negative correlation in magenta ($\alpha = .05$, corrected family-wise across the whole brain). Activation is plotted on the mean participant T1 image, and white dashed lines on the axial slice indicates the locations of the sagittal and coronal slices.

negative correlations with tracking load in the left supramarginal gyrus and angular gyri bilaterally, regions implicated in spatial attention, speech comprehension, and audiovisual integration (see: Seghier, 2013).

Speech × Task Interaction

Given our hypothesis that increasing attentional load would interfere with speech perception, we constrained our interaction analyses to speech-sensitive regions, both to aid in interpretation, and to reduce the inflation of our type II error rate by conservative correction for multiple comparisons across the whole brain. To achieve this, we constructed a speech perception mask using a union of the binarized masks for the main effect of speech type, and the speech × attention interaction contrasts, from Wild et al. (2012b), both thresholded at $\alpha = .05$, corrected for multiple comparisons across the whole brain (Worsley & Friston, 1995). This combination of masks: 1) allows us to focus our analyses on regions that are generally sensitive to speech quality, or are sensitive to speech quality depending on the attentional state; 2) provides a large (> 10,000 voxel) area within which to search; and 3) uses data from an independent cohort, preventing any dependencies between our mask and analyses (Kriegeskorte, Simmons, Bellgowan & Baker, 2009).

We observed a significant interaction between Task (Attend-Speech and Attend-MOT) and Speech Type (Clear, NV12, and NV6) in anterior insulae bilaterally, consistent with Wild et al. (2012b; see Figure 7). Two peaks within the left anterior insula were within the effective smoothing of our preprocessing (10mm apart, with an effective smoothing > 13mm), and so we averaged the parameter estimates across these peaks. To compare the response profiles across hemispheres, we ran a Region × Speech × Task mixed ANOVA on the parameter estimates from these regions. We found neither a main effect of Region ($F(1, 46) = 1.35$, $p = .25$, partial eta-squared = .03), nor any interactions between Region and our experimental conditions (all ps < .265). Accordingly, to simplify presentation, we averaged the parameter estimates across hemispheres.

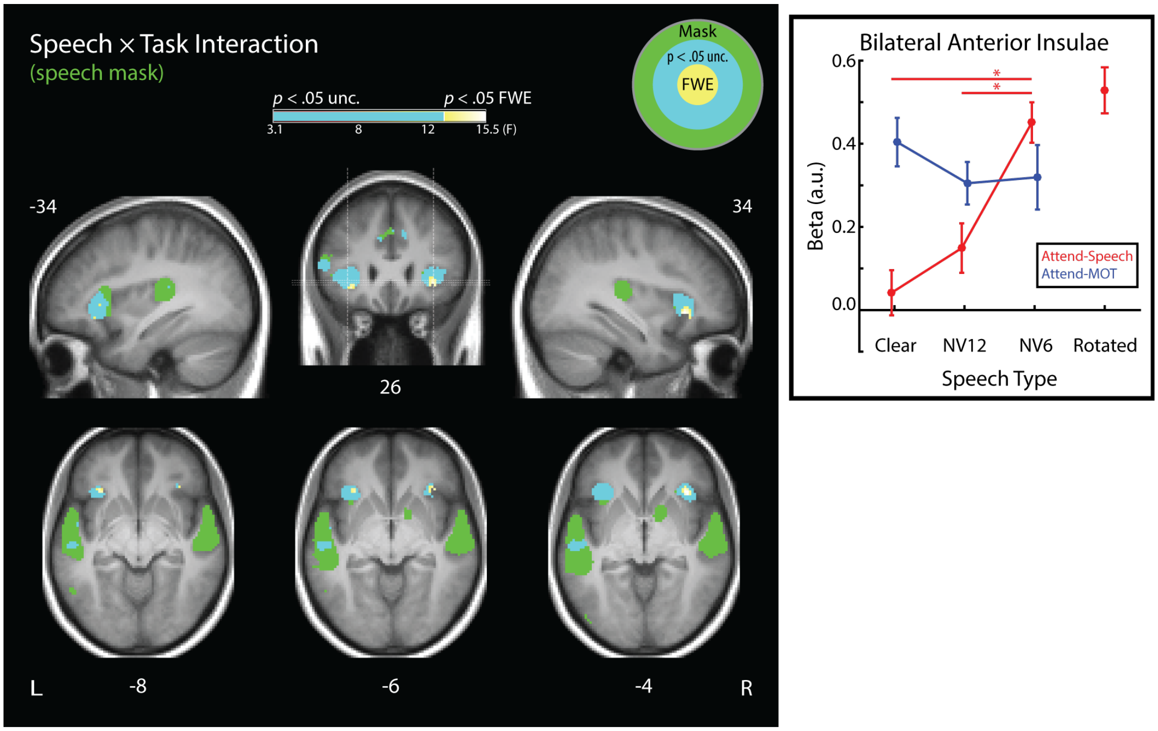

Figure 8. *Speech* × *Task interaction*. Left: Analyses were performed within a mask indicating cortex sensitive to speech. This mask was created by taking the union of the main effect of speech (Fig 4) and speech × attention interaction (Fig 6; equivalent to our speech × task contrast) maps (thresholded at α = .05, corrected family-wise across the whole brain) from Wild et al. (2012b). Cyan voxels exhibited an interaction between Speech Type and Task at an uncorrected threshold for visualization purposes (α = .05, uncorrected). Voxels that exhibited a significant interaction at a corrected threshold are indicated with a heat map correspond to their F-statistic (α = .05, corrected family-wise within the speech mask). Activation is plotted on the mean participant T1 image, and white dashed lines on the coronal slice indicates the locations of the sagittal and axial slices. Right: Graph depicts parameter estimates extracted from peak coordinates in the bilateral anterior insulae at each level of Speech Type, separately for Attend-Speech (red lines) and Attend-MOT (blue lines). Horizontal lines and asterisks indicate significant simple effects of Speech Type within the Attend-Speech Task (*ps* < .001, Šidák-corrected (Šidák, 1967) for 6 comparisons); within the Attend-MOT task, the effect of Speech Type was not significant in these voxels. Errors bars indicate SEM adjusted for within-subject measurements (Morey, 2008).

In order to determine the simple effects driving the two-way interaction, we analyzed the insular signal with a Speech Type × Task repeated-measures ANOVA. Post-hoc analysis of the Speech Type × Task interaction revealed that activity for NV6 was significantly greater than for NV12 or Clear speech during Attend – Speech ($\text{NV6}_{\text{Attend-Speech}} > \text{Clear}_{\text{Attend-Speech}}$: $t(23) = 6.14$, $p < .001$; $\text{NV6}_{\text{Attend-Speech}} > \text{NV12}_{\text{Attend-Speech}}$: $t(23) = 5.50$, $p < .001$), but that activity did not differ by Speech Type under the Attend-MOT condition (all *ps* > .408). This pattern of degradation-elevated activation selective to Attend-Speech is consistent with the response profile observed in Wild et al. (2012b).

Unlike Wild et al., (2012) we found elevated activity in the anterior insulae during our distractor task, suggesting that the anterior insular response may not be selective to speech. Consistent with this view, we observed a positive correlation with tracking load in an 8mm search sphere centred on the right anterior insula peak (family-wise error corrected within the sphere; $t(23) = 4.79$, $p = .002$), and a marginally positive correlation with tracking load in the left anterior insula ($t(23) = 2.83$, $p = .087$). Given the dependence of the anterior insular response on the load of our distractor task, a plausible explanation for the difference between our response profile and that of Wild et al (2012) is that our distractor task was more difficult, on average, than the ones used in this previous experiment.

Speech × Task × Load Interaction

In the previous section, we established that processing of speech in some brain regions depends on whether speech is the focus of attention. The next question is whether the effect of attention is all-or-nothing, in which case the interaction would only depend on which task participants performed, and would not be additionally modulated by different levels of tracking load (i.e., a 2-way, but not 3-way interaction; see Figure 1A). Alternatively, speech processing may share capacity with other cognitive processes, such that the interaction between Speech Type and Task would also depend on tracking load (i.e., a 3-way interaction; see Figure 1, B and C). We first looked for regions where the relationship between tracking load and BOLD activation depended on both Speech Type and Task, and then characterized responses in each of these regions by examining how the difference in activation level between full attention and distraction depended on both Speech Type and on Tracking Load. We predicted that the BOLD response would correlate inversely with MOT load during either clear or degraded speech trials, with the correlation during NV12 trials more similar to either Clear or NV6 speech.

During Attend-Speech, the total number of (task-irrelevant) dots on the screen changed in order to match the number of onscreen dots in the corresponding level of tracking load during Attend-MOT. Thus, the “tracking load” factor is present for both levels of Task, although cognitively it is very different for the two tasks, of course. Given the strong linear dependence of behavioural tracking performance on tracking load ($r_{median} = .94$), as well as our observation of widespread load-correlated activity during MOT performance, we decided to model tracking load as a parametric modulator on the columns modelling each of the speech types (separately for the two attentional tasks). The three-way interaction manifests as significant effects of both Task and Speech Type on the slope of the relationship between BOLD activity and Tracking Load.

As in our Speech Type × Task analysis, we examined the Speech Type × Task × Load interaction using the independently defined mask developed using two contrasts from Wild et al. (2012) described earlier. We observed a significant interaction in anterior areas on the bilateral superior temporal gyri (aSTG; see Figure 8; graph portrays parameter estimates extracted from each level of tracking load). We entered the extracted parameter estimates from the peak aSTG voxel in each hemisphere into a Region × Speech Type × Task mixed ANOVA. We did not find a main effect of region: $F(1, 46) = .094$, $p = .761$, partial eta-squared = .01; or any interactions involving Region and our experimental conditions (all ps > .528). Accordingly, we averaged the parameter estimates across these regions to produce a single aSTG response, in order to simplify presentation.

A two-way (Speech Type × Task) repeated-measures ANOVA on the extracted aSTG parameter estimates revealed that Speech Type markedly affected the weights for the load parametric modulator (i.e., slopes relating BOLD to attentional load magnitude) during Attend-MOT, but not during Attend-Speech. During Attend-MOT, Clear speech had a more negative load-dependent slope than either NV12 ($t(23) = -4.85$, $p < .001$) or NV6 ($t(23) = -2.87$, $p = .025$). We tested all six slopes for the three Speech Types × two tasks against 0 (no significant relationship between load and BOLD) using one-sample t-tests. Only Clear speech, during tracking, exhibited a significant (or even marginal) relationship (Šidák-corrected for 6 comparisons: $\text{Clear}_{\text{Attend-MOT}}$: $t(23) = -4.17$, $p = .002$), and this condition also exhibited a slope that was significantly different from those in all other conditions. Thus, the aSTG exhibited a three-way interaction and this was characterized by BOLD signal decreasing as load increased, but only for Clear speech, and only when attention was on the MOT task.

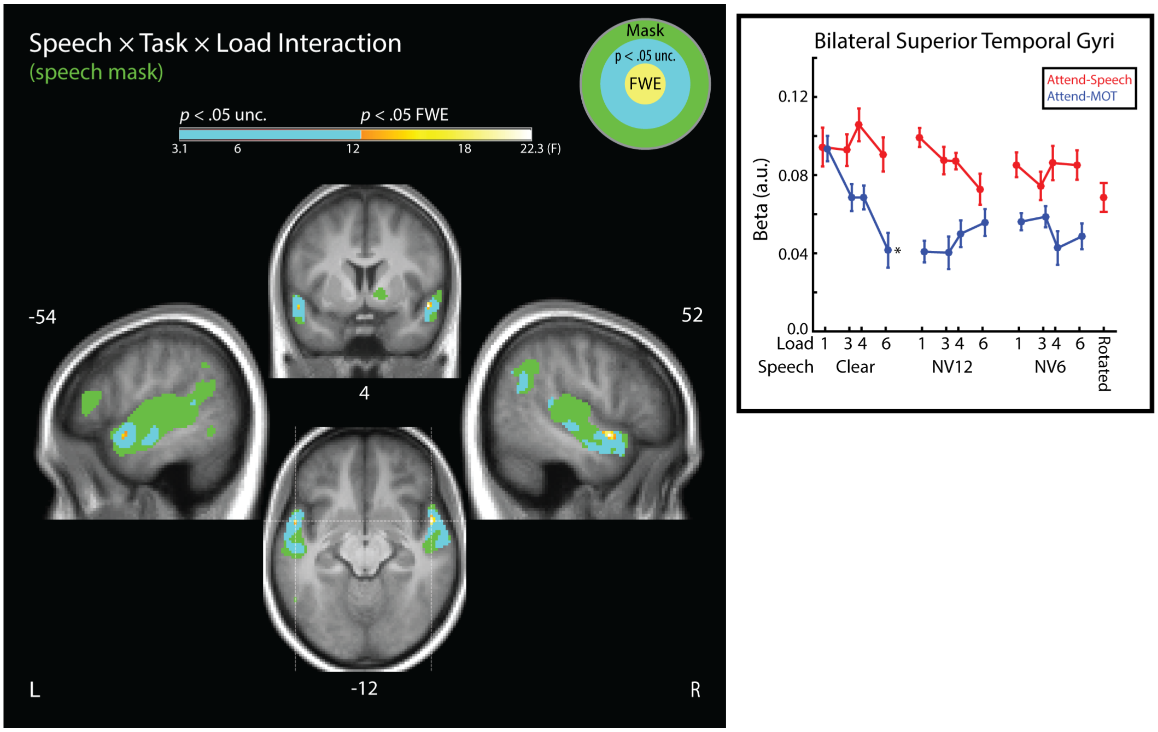

Figure 9. *Speech* × *Task* × *Load Interaction*. Left: Analyses were performed within the same mask used in Fig 7. In Cyan voxels, the slope relating BOLD activation to tracking load depended on both Task and Speech Type (α = .05, uncorrected). Voxels that exhibited a significant interaction at a corrected threshold are indicated with a heat map corresponding to their F-statistic (α = .05, corrected family-wise within the speech mask). Activation is plotted on the mean participant T1 image, and white dashed lines on the axial slice indicates the locations of the sagittal and coronal slices. Right: Graph depicts parameter estimates extracted from peak coordinates in the bilateral anterior superior temporal gyri at each level of Load and Speech Type, plotted separately for Attend-Speech (red) and Attend-MOT (blue). During Attend-MOT, the correlation between BOLD and tracking load was more negative for Clear speech than either NV speech. During Attend-Speech, there were no differences between speech types. Asterisk indicates the only significant correlation with tracking load (*p* = .002, Šidák-corrected for 6 comparisons). Errors bars indicate SEM adjusted for within-subject measurements (Morey, 2008).

We also analyzed the Speech × Task × Load interaction in these aSTG peak coordinates, treating each load level as a separate condition, in order to understand how the Speech by Task interaction changes as a function of tracking load during the MOT task. (see Figure 8). Since we did not expect tracking load to affect BOLD during Attend-Speech, and indeed we did not observe any such effect, we averaged over levels of tracking load during Attend-Speech, separately for each speech type. We then, for each Speech Type, subtracted the response during Attend-MOT at each level of Load from the overall response during Attend-Speech for that Speech Type. For this measure, zero indicated the same response for a given speech type during Attend-Speech and Attend-

MOT, and negative values indicated a weaker response for that Speech Type during Attend-MOT compared to during Attend-Speech.

We first tested whether the peak aSTG responses differed between Attend-Speech and Attend-MOT (i.e., one-sample t-tests against 0) for each Speech Type, at each level of Load. Clear, at the lowest level of Load (1 dot), was the only condition that elicited statistically indistinguishable levels of activity between Attend-Speech and Attend-MOT states (Attend-Speech ($t(23) = -0.31$, $p = .76$ uncorrected, $p = 1.00$ Šidák-corrected). Looking within the lowest level of tracking load, we found that the difference between Attend-Speech and Attend-MOT was significantly larger for NV12 ($t(23) = 5.61$, $p < .001$) and NV6 ($t(23) = 3.11$, $p = .015$) than it was for Clear. In other words, at the weakest level of load, Clear speech elicits a more similar response between tasks than was the case for NV12 or NV6, and is not significantly different from Attend-Speech. At higher levels of load, the difference between tasks did not depend on Speech Type (all *p*s > .193). Importantly, the significant decrease in activation for NV6 and NV12 between Attend-Speech and Attend-MOT at the least challenging level of MOT did not get any larger at more challenging levels, suggesting that activation in this region evoked by degraded speech is already effectively at floor when even a mild (1-dot MOT) distractor task is used.

In conclusion, we found that in the aSTG, the processing of highly intelligible degraded speech (both NV6 and NV12) depended markedly on which task participants performed; falling to very low values even at the least demanding level of the MOT task. In contrast, the processing of clear speech was similar when it was attended, and when it was heard while participants performed the least demanding MOT task, with activation falling steadily at more challenging levels of MOT. This imaging dissociation between NV12 and clear speech is interesting, given their similar intelligibilities.

Prefrontal Speech × Task × Load Interaction

There is a large body of work on the role of the left IFG (including ‘Broca’s Area’) in speech-related attentional control (Thompson-Schill, D’Esposito, Aguirre, & Farah, 1997;

Poldrack et al., 1999; Gold & Buckner, 2002; Davis & Johnsrude, 2003, 2007; Rodd, Johnsrude, & Davis, 2012; Wild, Davis, & Johnsrude, 2012), and the homologous right IFG has been implicated in domain-general attentional control (Corbetta & Shulman, 2002; Levy & Wagner, 2011; Hampshire et al., 2010; Fedorenko, Duncan, & Kanwisher, 2013), including exhibiting load-dependent activity during MOT (Tomasi, Ernst, Caparelli, & Chang, 2004; Tomasi, Wang, Wang, & Volkow, 2014). Both of these regions have also been implicated in dual-task interference (Herath et al, 2001; Jiang, Sae, & Kanwisher, 2004; Sabri, Humphries, Binder, & Liebenthal, 2013), including interference between auditory and visual tasks (Tombu et al, 2011; Wild et al., 2012; Finoia et al, 2015).

This previous work linking the IFG to attentional control in the context of speech, MOT, and dual-task paradigms motivated our investigation of load-dependent effects in bilateral IFG regions of interest. We constructed an anatomical mask using the LONI Probabilistic Atlas (LPBA40; Shattuck et al., 2007). This was a binary mask with voxels that were labelled as either the left or right IFG (RIFG), with a maximum likelihood threshold of 50%. This produced a mask with over 3800 voxels, and visual inspection confirmed that it provided good coverage of the IFG bilaterally.

Within this mask, we looked for voxels in which the relationship between tracking load and BOLD activity (i.e., beta weights on the parametric modulators modelling tracking load) depended on both Speech Type (3 levels) and attentional Task (2 levels) and found a significant peak (see Figure 9). To determine whether the pattern of the 3-way interaction was different from that in the bilateral aSTG, we ran a Region × Speech Type × Task mixed ANOVA on the extracted parameter estimates. Although we found a marginal Region × Task interaction ($F(1,46) = 2.99$, $p = .091$, partial eta-squared = .06), neither the main effect of Region, nor other interactions involving the factor Region were significant. The marginal interaction between region and task was driven by a larger difference between tasks in the RIFG than the aSTG.

As we did in the aSTG, we examined how the Speech Type by Task interaction changed as a function of tracking load during the MOT task. We averaged the peak RIFG

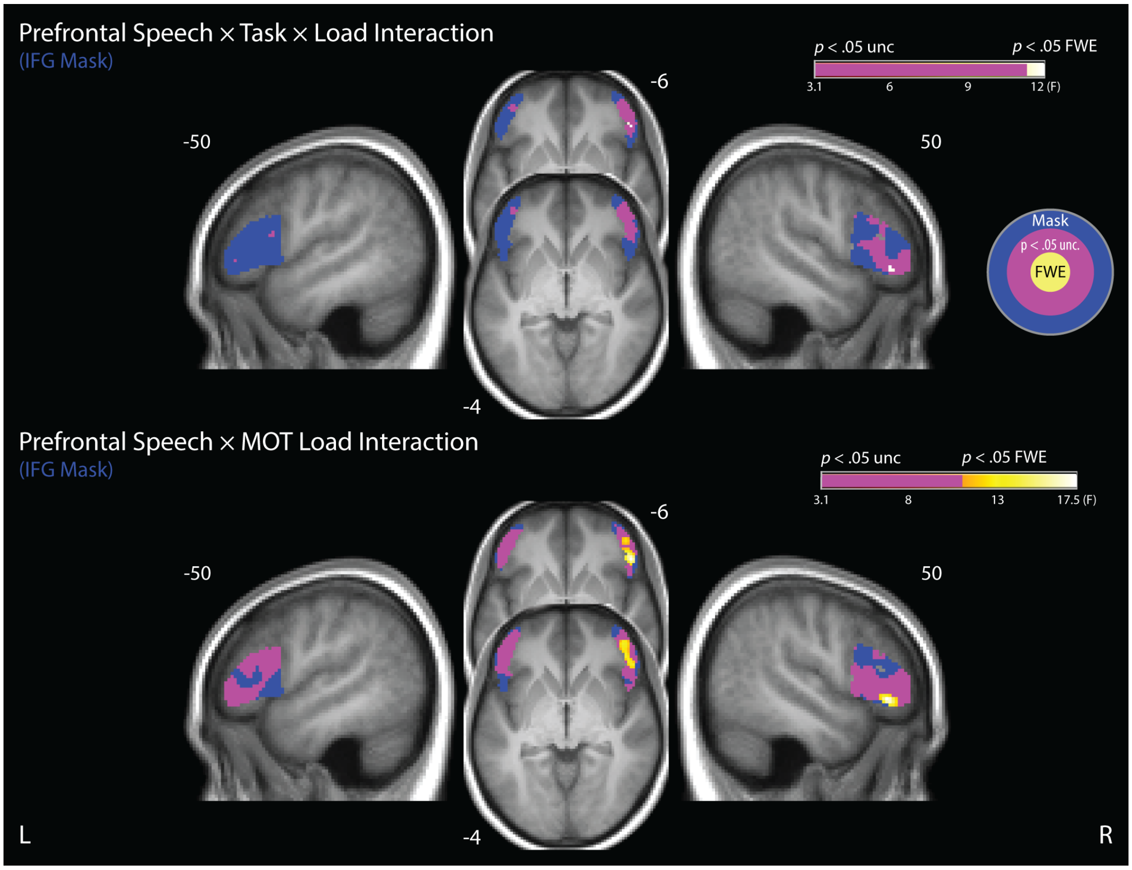

Figure 10. *Prefrontal Interactions*. Analyses were performed within a bilateral inferior frontal gyrus (IFG) region-of-interest defined using the LONI LPBA40 atlas (Shattuck et al., 2007), thresholded at a 50% maximum likelihood. Magenta voxels visualized where the slope relating BOLD activation and tracking load depended on both Task and Speech Type (α = .05, uncorrected). Voxels that exhibited the significant speech × task × load interaction at a corrected threshold (α = .05, corrected family-wise within the IFG mask) are indicated with a heat map corresponding to their F-statistic. Top: Coloured voxels indicate a significant interaction between Speech Type and Task on the slope of the relationship between BOLD activation and tracking load. Bottom: Coloured voxels corresponded to where slope of the relationship between BOLD activation and tracking load was different between Speech Types, based only on Attend-MOT trials. The activation patterns in peak voxels were not significantly different from those observed using the full model, even when comparing Attend-Speech trials across regions.

response over the tracking-load factor for each Speech Type when for Attend-Speech trials. For Attend-MOT trials, for each Speech Type, each level of load was treated separately. We ran a Region × Speech Type × Tracking Load mixed ANOVA on the differences between BOLD responses during Attend-MOT and the corresponding mean speech responses during Attend-Speech. We observed neither a significant main effect of Region nor interactions involving Region (*p*s > .350). As is the aSTG, the Speech Type × Load interaction was driven by the difference in activation between Attend-Speech and the lowest level of tracking load being statistically nonsignificant for Clear (one-sample t-tests against 0: $t(23) = 1.46$, $p = .158$ uncorrected, $p = .873$ Šidák-corrected), but significant for NV6 and NV12. At more challenging levels of tracking load, the difference between Attend-Speech and Attend-MOT was significant for both Clear speech and NV speech.

Given that 'tracking load' has cognitive reality only under Attend MOT, we conducted an exploratory analysis examining the simple two-way interaction between speech type and load only at the Attend-MOT level of Task (Figure 9, bottom). This contrast, which is necessarily more sensitive than the full model to speech differences unique to Attend-MOT, revealed two peaks in RIFG. The pattern of activity in these peaks was not significantly different to those observed in the RIFG (or aSTG) peaks using the full (three-way) interaction, suggesting that these regions are involved in similar speech processes.

Discussion

The core question that this experiment addressed was whether the perception of degraded speech depends on attention in an all-or-none fashion (i.e., speech perception exhibits a processing bottleneck), or whether it can be enhanced by the partial availability of attention (i.e., speech perception exhibits capacity-sharing). We searched for speech processing bottlenecks by looking for behavioural and neural correlates of speech perception (i.e., differential neural responses to clear and degraded speech) that were only present when participants focused on speech and that were eliminated even at the weakest, least challenging, level of the distractor MOT task. We searched for speech processes that share capacity by looking for correlates of speech perception that were modulated by the demands of the distractor task. We found evidence for speech regions that exhibit both bottleneck and capacity-sharing response profiles.

Our pattern of posttest recognition results was similar to Wild et al. (2012). As in Wild et al. (2012), participants were able to remember clear and highly intelligible degraded sentences better than chance, even when they were distracted, suggesting that for such stimuli speech perception is not entirely dependent on participants' attentional state. Memory was worse for sentences heard while participants were distracted, compared to when they were attending to speech, and distraction was more deleterious for degraded speech than for clear speech, even for degraded speech that was 100% intelligible.

Unlike Wild et al., (2012), who observed similar recognition scores for clear speech, regardless of the focus of attention, we found that memory was always poorer for sentences heard during Attend-MOT, relative to Attend-Speech, for all speech types. Our distractor task was probably more difficult, on average, than the one used in Wild et al. (2012), which may explain why it had a stronger effect on memory.

Although the focus of attention influenced the recognition of clear and degraded speech, we did not observe modulation of recognition scores under different MOT loads, despite finding neural correlates of load-dependent speech perception. This lack of modulation in memory scores may be due to our recognition test being a relatively insensitive measure of processing/encoding. There were only 9 observations at each level of MOT load, and the foils in the recognition test were entirely different from the targets – we cannot distinguish cases in which participants recognized all the words in a sentence (and so responded “old”) or only recognized one word (and responded the same way).

Interestingly, we found that recognition memory was better for NV12 speech than clear speech, despite being degraded. This finding appears to contrast with previous research that has documented poorer memory for degraded words than clear words (Rabbitt, 1966; Pichora-Fuller, Schneider, & Daneman, 1995; Surprenant, 1999; Murphy, Craik, Li, & Schneider, 2000). This memory enhancement for degraded speech relative to clear speech, observed uniquely when speech is the focus of attention, may reflect top-down influences over the perception of degraded speech that enhance encoding (c.f., Nairne, 1988; Hirshman, & Mulligan, 1991; Mulligan, 1996). In previous experiments that failed to find this effect, stimuli have either not had the contextual constraints of full sentences (Rabbitt, 1966; Surprenant, 1999; Murphy, Craik, Li, & Schneider, 2000), or have imposed segregation demands for noise- or speech-masking (Rabbitt, 1966; Pichora-Fuller, Schneider, & Daneman, 1995; Surprenant, 1999; Murphy, Craik, Li, & Schneider, 2000). These memory results suggest that regional increases in brain activity when attending to degraded, compared to clear, speech may also be involved in enhancing the encoding of degraded speech.

In the anterior insulae, there was increased activity for less intelligible speech, but only when participants focused on speech, consistent with Wild et al. (2012). Activity in this region was also correlated with tracking load during Attend-MOT, but not differently between speech types. We can further characterize the anterior insular response by examining whether responses to NV12 were more similar to clear speech (which are matched in intelligibility), or more similar to NV6 (since both are degraded, and the focus of attention mattered more for memory of both NV6 and NV12 stimuli compared to memory for clear speech). In this region, responses to NV12 were most similar to clear speech. This intelligibility-dependent response, in conjunction with the sensitivity to the demands of both tasks, may indicate that the anterior insulae have a role in monitoring the performance of whatever task is actively being attended to, consistent with proposals that it is involved in cognitive control (Duncan & Owen, 2000; Bunge et al., 2002**;** Fedorenko, Duncan, & Kanwisher, 2013; Shenhav, Botvinick, & Cohen, 2013; Cieslik et al, 2015) and, specifically, performance-monitoring (Wager et al., 2005**;** Dosenbach et al., 2006; Vaden et al, 2013; Lamichhane, Adhikari, & Dhamala, 2016). These results do not necessarily suggest that there is not attentionally enhanced processing for NV12 sentences in this region. Behaviourally, after all, memory for NV12 materials was better than for clear materials. The single volume we collect with our sparse acquisition cannot measure the timecourse of sentence perception. It may be that enhanced activity is more fleeting than the slow BOLD response can index (e.g., if the region responded more to NV12 speech than clear speech only at the beginning of the sentence, in order to establish a predictive context).

For clear speech, BOLD activity in the anterior STG and RIFG did not differ between full attention and the lowest level of MOT load, but was negatively correlated with MOT load. In contrast, BOLD activity for degraded speech in these regions dropped sharply between full attention and the lowest level of MOT load, and then did not differ among different levels of MOT load. This interaction between speech type and tracking load suggests that at some point during speech perception, there are processes that are affected by both the demands of speech perception and the demands of object tracking. Because the cognitive demands of both tasks interact, this is evidence that they share cognitive capacity.

The pattern of results in STG/RIFG suggest that there are multiple ways that attention can influence speech perception. Some speech processes exhibit a bottleneck, whereas others exhibiting capacity sharing. For clear speech, there is evidence for speech processes that depend on a shared capacity, since the STG/RIFG responses were negatively correlated with tracking load. The task- (but not load-) dependent responses to degraded speech, in contrast, suggest that processes involved in the comprehension of degraded speech critically require focused attention.

Consistent with the interpretation of anterior STS activity in Wild et al. (2012b), activity in STG/RIFG may reflect attentionally enhanced intelligibility, i.e., our STG activity may reflect intelligibility after knowledge-guided interpretive and repair processes have augmented the strict ‘bottom-up’ intelligibility of degraded speech. Such processes may be ‘gated’ by attention in an all-or-nothing fashion: the task-specific response in the anterior insulae makes this region a candidate for such a gating function, with the STG a recipient of this modulation. Indeed, primate anatomical experiments reveal that the ventrolateral prefrontal cortex is strongly interconnected with the rostral temporal lobe (Romanski et al., 1999).

In terms of the capacity-dependent activity seen for clear speech (i.e., the negative correlation with tracking load), it may be the case that other regions are modulating this STG response as well. If a region was directly involved in a capacity-dependent speech process, then we should expect to find a similar dependence on tracking load for all speech types, assuming that this processes is allocated cognitive capacity regardless of whether or not it is sufficient to allow for successful comprehension. The STG response to clear speech depends on attention, but in a different way to degraded speech. Whereas the processing of degraded speech in STG is apparently at floor as soon as attention is elsewhere, processing of clear speech in the absence of attention appears to depend on spare cognitive capacity. The source(s) of the modulation of this region is/are yet to be determined, but both the angular and supramarginal gyri exhibited activity that was negatively correlated with tracking load, and both appear to be involved in speech processing and multimodal integration (see: Seghier, 2013).

An attentionally dependent RIFG response was surprising, given previous observations that the LIFG has attentionally dependent role in speech perception (e.g., Sabri, 2008; Wild et al., 2012). The bilateral IFG have been observed to play a role in MOT (Tomasi, Ernst, Caparelli, & Chang, 2004; Tomasi, Wang, Wang, & Volkow, 2014), and so this response may be a byproduct of the task that we used to manipulate attention. The apparent lack of functional difference between the anterior STG and RIFG responses should be investigated in further experiments, in order to understand whether these regions have distinct roles in speech perception.

Attentionally dependent modulation of speech processing was observed in high-level auditory processing regions (Kaas, Hackett, & Tramo, 1999) that are responsive to linguistic features in humans (Davis & Johnsrude, 2003; Okada et al., 2010; Evans et al., 2014), regions that were also attentionally modulated in Wild et al. (2012b). Further experiments should more thoroughly examine the conditions under which attention influences different levels within the speech hierarchy.

The results of our experiment have implications for the diagnosis of clinically meaningful hearing impairment. The role of 'listening effort' in speech perception is of growing interest in the field of audiology (see: McGarrigle et al., 2014; Johnsrude & Rodd, 2016). The results of our experiment show that mild distraction can dramatically change neural responses to highly intelligible degraded speech. Traditional methods of hearing assessment involve audiometric testing of pure-tone perceptual thresholds across a range of frequencies, or measuring the minimum amplitude threshold for speech comprehension (American Speech-Language-Hearing Association). More ecologically valid measures of speech comprehension should involve naturalistic, contextualized speech (to allow for the top-down modulation we believe to depend on contextual constraint); testing individuals using degraded speech (to evoke attentional processes that we found to be essential for processing of degraded, but not clear, speech); and measuring individuals' perception of clear and degraded speech under a cognitive load (in order to estimate their ability to compensate for speech degradation). These methods may supplement traditional assessment methods in informative ways.

Conclusion

We have provided further evidence that speech perception is influenced by attention in the anterior superior temporal gyri, inferior frontal gyri, and anterior insulae (Wild et al., 2012). We have extended previous research by showing that whereas some regions in the system sensitive to speech exhibit attentionally gated processing (i.e., are processing bottlenecks), other regions appear to be able to share cognitive resources with visual attention. Furthermore, we have found that these speech processes enhance the perception of speech that, while acoustically degraded, is as intelligible as clear speech. Future studies should investigate how the attentionally dependent speech processes that we characterize in this experiment can contribute not only to theories of speech perception, but also to improvements in treatment and diagnostics for clinical populations.

References

Akeroyd, M. A. (2008). Are individual differences in speech reception related to individual differences in cognitive ability? A survey of twenty experimental studies with normal and hearing-impaired adults. *International Journal of Audiology*, *47*(2), S53–S71.

Allen, R., McGeorge, P., Pearson, D. G., & Milne, A. (2006). Multiple-target tracking: A role for working memory? *Quarterly Journal of Experimental Psychology*, *59*(6), 1101–16.

Alvarez, G. A., & Franconeri, S. L. (2007). How many objects can you track? Evidence for a resource-limited attentive tracking mechanism. *Journal of Vision*, *7*(13), 14.1–10.

Alvarez, G. A., Horowitz, T. S., Arsenio, H. C., DiMase, J. S., & Wolfe, J. M. (2005). Do multielement visual tracking and visual search draw continuously on the same visual attention resources? *Journal of Experimental Psychology: Human Perception and Performance*, *31*(4), 643.

American Speech-Language-Hearing Association. (1997). Guidelines for audiologic screening.

Amichetti, N. M., Stanley, R. S., White, A. G., & Wingfield, A. (2013). Monitoring the capacity of working memory: Executive control and effects of listening effort. *Memory & Cognition*, *41*(6), 839–849.

D'Agostino, E., Maes, F., Vandermeulen, D., & Suetens, P. (2004). Non-rigid atlas-to-image registration by minimization of class-conditional image entropy. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 745–753).

Besser, J., Koelewijn, T., Zekveld, A. A., Kramer, S. E., & Festen, J. M. (2013). How linguistic closure and verbal working memory relate to speech recognition in noise —A review. *Trends in Amplification*, *17*(2), 75–93.

Bettencourt, K. C., & Somers, D. C. (2009). Effects of target enhancement and distractor suppression on multiple object tracking capacity. *Journal of Vision*, *9*, 9.

Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex*, *19*(12), 2767–2796.

Broadbent, D.E. (1958). *Perception and communication*. London: Pergamon Press.

Bunge, S. A., Dudukovic, N. M., Thomason, M. E., Vaidya, C. J., & Gabrieli, J. D. E. (2002). Immature frontal lobe contributions to cognitive control in children: evidence from fMRI. *Neuron*, *33*(2), 301–311.

Burkholder, R. A., Pisoni, D. B., & Svirsky, M. A. (2005). Effects of a cochlear implant simulation on immediate memory in normal-hearing adults. *International Journal of Audiology*, *44*(10), 551–558.

Cavanagh, P., & Alvarez, G. A. (2005). Tracking multiple targets with multifocal attention. *Trends in Cognitive Sciences*, *9*(7), 349–354.

Cieslik, E. C., Mueller, V. I., Eickhoff, C. R., Langner, R., & Eickhoff, S. B. (2015). Three key regions for supervisory attentional control: evidence from neuroimaging meta-analyses. *Neuroscience & Biobehavioral Reviews*, *48*, 22–34.

Cowan, N. (2001). Metatheory of storage capacity limits. *Behavioral and Brain Sciences*, *24*(01), 154–176.

Culham, J. C., Brandt, S. A., Cavanagh, P., Kanwisher, N. G., Dale, A. M., & Tootell, R. B. H. (1998). Cortical fMRI activation produced by attentive tracking of moving targets. *Journal of Neurophysiology*, *80*(5), 2657–2670.

Culham, J. C., Cavanagh, P., & Kanwisher, N. G. (2001). Attention response functions: Characterizing brain areas using fMRI activation during parametric variations of attentional load. *Neuron*, *32*, 737–745.

Davis, M. H., Ford, M. a, Kherif, F., & Johnsrude, I. S. (2011). Does semantic context benefit speech understanding through “top-down” processes? Evidence from time-resolved sparse fMRI. *Journal of Cognitive Neuroscience*, *23*(12), 3914–32.

Davis, M. H., Johnsrude, I. S., Hervais-Adelman, A., Taylor, K., & McGettigan, C. (2005). Lexical information drives perceptual learning of distorted speech: evidence from the comprehension of noise-vocoded sentences. *Journal of Experimental Psychology: General*, *134*(2), 222.

Davis, M., & Johnsrude, I. (2003). Hierarchical processing in spoken language comprehension. *The Journal of Neuroscience*, *23*(8), 3423–3431.

Dosenbach, N. U. F., Visscher, K. M., Palmer, E. D., Miezin, F. M., Wenger, K. K., Kang, H. C., … Petersen, S. E. (2006). A core system for the implementation of task sets. *Neuron*, *50*(5), 799–812.

Downs, D. W. (1982). Effects of hearing aid use on speech discrimination and listening effort. *Journal of Speech and Hearing Disorders*, *47*(2), 189–193.

Duncan, J., & Owen, A. M. (2000). Common regions of the human frontal lobe recruited by diverse cognitive demands. *Trends in Neurosciences*, *23*(10), 475–483.

Evans, S., Kyong, J. S., Rosen, S., Golestani, N., Warren, J. E., McGettigan, C., … Scott, S. K. (2014). The pathways for intelligible speech: multivariate and univariate perspectives. *Cerebral Cortex*, *24*(9), 2350–2361.

Fedorenko, E., Duncan, J., & Kanwisher, N. (2013). Broad domain generality in focal regions of frontal and parietal cortex. *Proceedings of the National Academy of Sciences*, *110*(41), 16616–21.

Fonov, V. S., Evans, A. C., McKinstry, R. C., Almli, C. R., & Collins, D. L. (2009). Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *NeuroImage*, *47*, S102.

Feuerstein, J. F. (1992). Monaural versus binaural hearing: ease of listening, word recognition, and attentional effort. *Ear and Hearing*, *13*(2), 80–86.

Finoia, P., Mitchell, D. J., Hauk, O., Beste, C., Pizzella, V., & Duncan, J. (2015). Concurrent brain responses to separate auditory and visual targets. *Journal of Neurophysiology*, *114*(2), 1239–1247.

Francis, A. L., & Nusbaum, H. C. (2009). Effects of intelligibility on working memory demand for speech perception. *Attention, Perception, & Psychophysics*, *71*(6), 1360–1374.

Franconeri, S. L., Alvarez, G. A., & Cavanagh, P. (2013). Flexible cognitive resources: Competitive content maps for attention and memory. *Trends in Cognitive Sciences*, *17*(3), 134–141.

Giraud, A. L., Kell, C., Thierfelder, C., Sterzer, P., Russ, M. O., Preibisch, C., & Kleinschmidt, A. (2004). Contributions of sensory input, auditory search and verbal comprehension to cortical activity during speech processing. *Cerebral Cortex*, *14*(3), 247–255.

Gold, B. T., & Buckner, R. L. (2002). Common prefrontal regions coactivate with dissociable posterior regions during controlled semantic and phonological tasks. *Neuron*, *35*(4), 803–812.

Gosselin, P. A., & Gagné, J. P. (2010). Use of a Dual-Task Paradigm to Measure Listening Effort. *CJSLPA*, *34*(1), 43-51.

Greenwood, D. D. (1990). A cochlear frequency-position function for several species—29 years later. *The Journal of the Acoustical Society of America*, *87*(6), 2592–2605.

Hall, D. A., Haggard, M. P., Akeroyd, M. A., Palmer, A. R., Summerfield, A. Q., Elliott, M. R., … Bowtell, R. W. (1999). Sparse temporal sampling in auditory fMRI. *Human Brain Mapping*, *7*(3), 213–223.

Hackett, T. A., Stepniewska, I., & Kaas, J. H. (1999). Prefrontal connections of the parabelt auditory cortex in macaque monkeys. *Brain Research*, *817*(1), 45–58.

Hampshire, A., Chamberlain, S. R., Monti, M. M., Duncan, J., & Owen, A. M. (2010). The role of the right inferior frontal gyrus: inhibition and attentional control. *NeuroImage*, *50*(3), 1313–1319.

Heald, S. L., & Nusbaum, H. C. (2014). Speech perception as an active cognitive process. *Frontiers in Systems Neuroscience*, *8*(March), 35. doi:10.3389/fnsys.2014.00035

Henson, R. N. A., & Penny, W. D. (2003). ANOVAs and SPM. *Wellcome Department of Imaging Neuroscience, London, UK*.

Herath, P., Klingberg, T., Young, J., Amunts, K., & Roland, P. (2001). Neural correlates of dual task interference can be dissociated from those of divided attention: an fMRI study. *Cerebral Cortex*, *11*(9), 796–805.

Hervais-Adelman, A. G., Carlyon, R. P., Johnsrude, I. S., & Davis, M. H. (2012). Brain regions recruited for the effortful comprehension of noise-vocoded words. *Language and Cognitive Processes*, *27*(7-8), 1145–1166.

Hirshman, E., & Mulligan, N. (1991). Perceptual interference improves explicit memory but does not enhance data-driven processing. Journal of Experimental Psychology: Learning, Memory, and Cognition, 17(3), 507.

Holcombe, A. O., & Chen, W. Y. (2012). Exhausting attentional tracking resources with a single fast-moving object. *Cognition*, *123*(2), 218–228.

Howe, P. D. L., Cohen, M. A., Pinto, Y., & Horowitz, T. S. (2010). Distinguishing between parallel and serial accounts of multiple object tracking. *Journal of Vision*, *10*(8), 11.

Howe, P. D., Horowitz, T. S., Morocz, I. A., Wolfe, J., & Livingstone, M. S. (2009). Using fMRI to distinguish components of the multiple object tracking task. *Journal of Vision*, *9*(4), 10.1–11.

Johnsrude, I. S., & Rodd, J. M. (2016). Chapter 40 - Factors That Increase Processing Demands When Listening to Speech. In G. Hickok & S. L. Small (Eds.), *Neurobiology of Language* (pp. 491–502). San Diego: Academic Press.

Jovicich, J., Peters, R. J., Koch, C., Braun, J., Chang, L., & Ernst, T. (2001). Brain areas specific for attentional load in a motion-tracking task. *Journal of Cognitive Neuroscience*, *13*(8), 1048–1058.

Kaas, J. H., Hackett, T. A., & Tramo, M. J. (1999). Auditory processing in primate cerebral cortex. *Current opinion in neurobiology*, *9*(2), 164-170.

Kahneman, D. (1973). *Attention and Effort*. Englewood Cliffs, N.J.: Prentice-Hall.

Kirchner, W. K. (1958). Age differences in short-term retention of rapidly changing information. *Journal of experimental psychology*, *55*(4), 352-358.

Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S. F., & Baker, C. I. (2009). Circular analysis in systems neuroscience: the dangers of double dipping. *Nature Neuroscience*, *12*(5), 535–540.

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

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, *74*(6), 431-461.

Luce, P., Feustel, T., & Pisoni, D. (1983). Capacity Demands in Short-Term Memory for Synthetic and Natural Speech. *Human Factors*, *25*(1), 17–32.

McGarrigle, R., Munro, K. J., Dawes, P., Stewart, A. J., Moore, D. R., Barry, J. G., & Amitay, S. (2014). Listening effort and fatigue: What exactly are we measuring? A British Society of Audiology Cognition in Hearing Special Interest Group ‘white paper’. *International journal of audiology*.

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorial in Quantitative Methods for Psychology*, *4*(2), 61–64.

Mulligan, N. W. (1996). The effects of perceptual interference at encoding on implicit memory, explicit memory, and memory for source. Journal of Experimental Psychology: Learning, Memory, and Cognition, 22(5), 1067.

Murphy, D. R., Craik, F. I., Li, K. Z., & Schneider, B. A. (2000). Comparing the effects of aging and background noise of short-term memory performance. Psychology and aging, 15(2), 323.

Nairne, J. S. (1988). The mnemonic value of perceptual identification. Journal of Experimental Psychology: Learning, Memory, and Cognition, 14(2), 248.

Navon, D., & Miller, J. (2002). Queuing or sharing? A critical evaluation of the single-bottleneck notion. *Cognitive Psychology*, *44*(3), 193–251.

Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review*, *115*(2), 357–95.

Norris, D., McQueen, J. M., & Cutler, A. (2016). Prediction, Bayesian inference and feedback in speech recognition. *Language, Cognition and Neuroscience*, *31*(1), 4–18.

Obleser, J., Eisner, F., & Kotz, S. A. (2008). Bilateral speech comprehension reflects differential sensitivity to spectral and temporal features. *The Journal of Neuroscience*, *28*(32), 8116–8123.

Obleser, J., & Kotz, S. A. (2010). Expectancy constraints in degraded speech modulate the language comprehension network. *Cerebral Cortex*, *20*(3), 633–40.

Obleser, J., Wise, R. J. S., Alex Dresner, M., & Scott, S. K. (2007). Functional integration across brain regions improves speech perception under adverse listening conditions. *The Journal of Neuroscience*, *27*(9), 2283–9.

Obleser, J., Wöstmann, M., Hellbernd, N., Wilsch, A., & Maess, B. (2012). Adverse listening conditions and memory load drive a common alpha oscillatory network. *The Journal of Neuroscience*, *32*(36), 12376–12383.

Okada, K., Rong, F., Venezia, J., Matchin, W., Hsieh, I. H., Saberi, K., … Hickok, G. (2010). Hierarchical organization of human auditory cortex: Evidence from acoustic invariance in the response to intelligible speech. *Cerebral Cortex*, *20*(10), 2486–2495.

Pals, C., Sarampalis, A., & Bacskent, D. (2013). Listening effort with cochlear implant simulations. *Journal of Speech, Language, and Hearing Research*, *56*(4), 1075–1084.

Pashler, H. (1984). Processing stages in overlapping tasks: evidence for a central bottleneck. *Journal of Experimental Psychology: Human Perception and Performance*, *10*(3), 358-377.

Pashler, H. (1994). Dual-task interference in simple tasks: data and theory. *Psychological Bulletin*, *116*(2), 220–244.

Pashler, H. & Sutherland, S. (1998). *The Psychology of Attention* (Vol. 15). MIT press Cambridge, MA.

Pichora-Fuller, M. K., Schneider, B. A., & Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *The Journal of the Acoustical Society of America*, *97*(1), 593–608.

Piquado, T., Cousins, K. a Q., Wingfield, A., & Miller, P. (2010). Effects of degraded sensory input on memory for speech: behavioral data and a test of biologically constrained computational models. *Brain Research*, *1365*, 48–65.

Poldrack, R. A., Wagner, A. D., Prull, M. W., Desmond, J. E., Glover, G. H., & Gabrieli, J. D. E. (1999). Functional specialization for semantic and phonological processing in the left inferior prefrontal cortex. *Neuroimage*, *10*(1), 15–35.

Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: evidence for a parallel tracking mechanism. *Spatial Vision*, *3*(3), 179–197.

Rabbitt, P. (1966). Recognition: Memory for words correctly heard in noise. Psychonomic Science, 6(8), 383-384.

Rabbitt, P. M. (1968). Channel-capacity, intelligibility and immediate memory. *The Quarterly Journal of Experimental Psychology*, *20*(3), 241–248.

Rodd, J. M., Davis, M. H., & Johnsrude, I. S. (2005). The neural mechanisms of speech comprehension: fMRI studies of semantic ambiguity. *Cerebral Cortex*, *15*(8), 1261-1269.

Romanski, L. M., Bates, J. F., & Goldman-Rakic, P. S. (1999). Auditory belt and parabelt projections to the prefrontal cortex in the rhesus monkey. *Journal of Comparative Neurology*, *403*(2), 141–157.

Sabri, M., Binder, J. R., Desai, R., Medler, D. A., Leitl, M. D., & Liebenthal, E. (2008). Attentional and linguistic interactions in speech perception. *NeuroImage*, *39*(3), 1444–1456.

Sabri, M., Humphries, C., Binder, J. R., & Liebenthal, E. (2013). Neural events leading to and associated with detection of sounds under high processing load. *Human Brain Mapping*, *34*(3), 587–597.

Scholl, B. J. (2009). What Have We Learned about Attention from Multiple-Object Tracking (and Vice Versa)? *Computation Cognition and Pylyshyn*, (2), 49–78.

Scott, S. K., Blank, C. C., Rosen, S., & Wise, R. J. S. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain*, *123*(12), 2400–2406.

Seghier, M. (2012). The Angular Gyrus: Multiple Functions and Multiple Subdivisions. *The Neuroscientist*, *19*(1), 43–61.

Shannon, R., Zeng, F., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science*, *270*(5234), 303–304.

Shattuck, D. W., Mirza, M., Adisetiyo, V., Hojatkashani, C., Salamon, G., Narr, K. L., … Toga, A. W. (2008). Construction of a 3D probabilistic atlas of human cortical structures. *Neuroimage*, *39*(3), 1064–1080.

Shenhav, A., Botvinick, M. M., & Cohen, J. D. (2013). The expected value of control: an integrative theory of anterior cingulate cortex function. *Neuron*, *79*(2), 217–40.

Surprenant, A. M. (1999). The effect of noise on memory for spoken syllables. *International Journal of Psychology*, *34*(5-6), 328-333.

Humes, L. E., Kidd, G. R., & Lentz, J. J. (2013). Auditory and cognitive factors underlying individual differences in aided speech-understanding among older adults. *Frontiers in Systems Neuroscience*.

Šidák, Z. (1967). Rectangular confidence regions for the means of multivariate normal distributions. *Journal of the American Statistical Association*, *62*(318), 626–633.

Skipper, J. I., Nusbaum, H. C., & Small, S. L. (2005). Listening to talking faces: motor cortical activation during speech perception. *Neuroimage*, *25*(1), 76–89.

Thompson-Schill, S. L., D’Esposito, M., Aguirre, G. K., & Farah, M. J. (1997). Role of left inferior prefrontal cortex in retrieval of semantic knowledge: A reevaluation. *Proceedings of the National Academy of Sciences*, *94*(26), 14792–14797.

Tomasi, D., Wang, R., Wang, G. J., & Volkow, N. D. (2014). Functional connectivity and brain activation: a synergistic approach. *Cerebral Cortex*, *24*(10), 2619–2629.

Tomasi, D., Wang, R., Wang, G.-J., Volkow, N. D., Culham, J. C., Cavanagh, P., … Chang, L. (2001). Practice-induced changes of brain function during visual attention: A parametric fMRI study at 4 Tesla. *Cerebral Cortex*, *23*(4), 1414–1421.

Tombu, M. N., Asplund, C. L., Dux, P. E., Godwin, D., Martin, J. W., & Marois, R. (2011). A unified attentional bottleneck in the human brain. *Proceedings of the National Academy of Sciences*, *108*(33), 13426–13431.

Tombu, M., & Jolicœur, P. (2003). A central capacity sharing model of dual-task performance. *Journal of Experimental Psychology: Human Perception and Performance*, *29*(1), 3-18.

Tombu, M., & Jolicœur, P. (2005). Testing the predictions of the central capacity sharing model. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(4), 790-802.

Tombu, M., & Seiffert, A. E. (2008). Attentional costs in multiple-object tracking. *Cognition*, *108*(1), 1–25.

Vaden, K. I., Kuchinsky, S. E., Cute, S. L., Ahlstrom, J. B., Dubno, J. R., & Eckert, M. A. (2013). The cingulo-opercular network provides word-recognition benefit. *The Journal of Neuroscience*, *33*(48), 18979–18986.

Wager, T. D., Sylvester, C.Y., Lacey, S. C., Nee, D. E., Franklin, M., & Jonides, J. (2005). Common and unique components of response inhibition revealed by fMRI. *Neuroimage*, *27*(2), 323–340.

Wendorf, C. A. (2004). Primer on multiple regression coding: Common forms and the additional case of repeated contrasts. *Understanding Statistics*, *3*(1), 47–57.

Wild, C.J., (2012). Predictive Coding: How the Human Brain Uses Context to Facilitate the Perception of Degraded Speech. Retrieved from http://hdl.handle.net/1974/7511.

Wild, C. J., Davis, M. H., & Johnsrude, I. S. (2012a). Human auditory cortex is sensitive to the perceived clarity of speech. *NeuroImage*, *60*(2), 1490–502.

Wild, C. J., Yusuf, A., Wilson, D. E., Peelle, J. E., Davis, M. H., & Johnsrude, I. S. (2012b). Effortful listening: the processing of degraded speech depends critically on attention. *The Journal of Neuroscience*, *32*(40), 14010–21.

Wilson, M. D. (1988). The MRC Psycholinguistic Database: Machine Readable Dictionary, Version 2. Behavioural Research Methods, Instruments and Computers 20 (1): 6--11. *World Health Organization. The ICD-10 Classification of Mental and Behavioural Disorders. Geneva: World Health Organization*.

Wingfield, A., Tun, P. A., & McCoy, S. L. (2005). Hearing loss in older adulthood what it is and how it interacts with cognitive performance. *Current Directions in Psychological Science*, *14*(3), 144–148.

Worsley, K. J., & Friston, K. J. (1995). Analysis of fMRI time-series revisited—again. *Neuroimage*, *2*(3), 173-181.

Appendix A: Sentence Materials

his handwriting was very difficult to read
thunder was heard when the children were all in their rooms
she loved stories about fairies wizards and dragons
their holiday was quite short and would end soon
the lawyer has quite a large salary
a game of chess can last for four hours
trains are often delayed by bad weather
the woman was hoping to discover the name and address of the culprit
she grew tomatoes in her greenhouse
they drove from the seaside to the city at the end of the day
the juice was served in a large jug
he guessed the answer to the question in the exam
the television program was a success
the queen went on a tour of the country that summer
the blunt knife was rather awkward to use
the soldier had a map that showed him all the details
daisies will begin to grow quite soon
there were many sparrows in the sky just above the trees
the neighbors made a lot of noise last night
the furniture in the dining room was removed when the room was decorated
a spoon was used to stir the cup of tea
his new clothes were from France
the kettle had some water in it
the top of the tower had a wonderful view of the city
there were bracelets and necklaces in her jewellery box
there was a really beautiful sunset that evening
it is common for people to avoid the dentist
the view from the top of the ridge was amazing
he enjoyed the beauty of the hills
the bruise on his knee was quite painful
the elephant was huge just as the circus had wanted
the bride smiled at the photo of her wedding
the child was sad when her toys were damaged
she wrote her secrets in her diary
the bishop was welcomed into the chapel
the town had pubs that were quite cheap and easy to find
the boat drifted across the pond
the mayor used cash to bribe the reporters before they exposed him to the public
her daughter was too young for the disco
the beef was rare just as the customer had requested
the lecturer insisted that the students should submit their essays on time
his girlfriend had chosen the picture on the wall
the shoes were not the colour that the young girl wanted

the crooked tree was in danger
it was the crew that remained when the final lifeboat left the ship
the chocolates and the flowers were bought from the nearest florist
the truce was broken when more guns were delivered
the group of friends got a taxi home after they left the nightclub
an angry crowd was turned back at the government building
there were mice in the cave
they thought that the house was haunted
the audience was quiet when the song was started
the rice was cooked in a large saucepan
it was too cold to go camping in winter
the artefacts found at the dig were made of bronze
the thief started to sprint very fast
the drink was too hot for the baby
he ironed his shirt before he wore it
the safety rules of the apartment were important to follow
it is best if the hamster stays in the shade during the summer
the win helped our team advance to the play-offs
they were concerned when the kid laughed at violent movies
they hoped that the pill did not have any side effects
he explained that the arch had been built by the Romans
the pantry contained ingredients he had never seen before
the pole did not support their weight as they climbed over the gate
the canyon was filled with haze on sunny days
it was the women that complained when the old bingo hall was closed
the track turned north towards the forest
everyone was worried as the exam was much harder than expected
the shrubs are watered regularly by the gardener
the public stopped attending the games after a bad start to the season
roses will start to bloom very soon
a splash of gin tastes really good with ice and lemon
the vessel was still watertight even when badly battered
the drought was eased by the arrival of the monsoon
there has been a tree towering above this house for the last fifty years
it is because the ant lived under the rocks that it survived the explosion
the gambler lost most of his money at the races
taking a nap can help you stay up later
it was a cloudy week so the residents stayed in their dormitories
he searched the pack for the ace of hearts
the old house was for sale
the children thought the dolphin was beautiful
the tray should have been returned to the kitchen
we noticed that the pen shook when the man signed the form
she claimed that the bran tasted much nicer
the tie attracted attention because of its odd appearance

the plane flew over the buildings
there were books in the cellar
the croquet game could begin after the lawn was mowed
she thought her jacket made her look very smart
the bait should be suitable for catching rats
the pain tempted him to abort the climb
the fumes from the factory are unbearable in the village
a severe storm left the walnut tree badly damaged
the author wrote the book that year
the garage was closed on weekends
it was unfortunate that the fog was so thick
they thought that the stable would cost more than the house to heat
the coin was thrown onto the floor
it was obvious that the junction was dangerous to drive around
the den should be an ideal place to study
the dentist needed somewhere to relax at the end of the day
opening the can takes a long time with a rusty penknife
aeroplanes are currently the best way to travel
the rowing team veered into the bank at the start of the race
the platform started creaking alarmingly during the speech
the dock should be fairly quiet on saturdays
the gems found in the store were not worth very much money
taking a hostage allowed the robbers to make their escape
some milk was borrowed from his neighbour
the carpet and the curtains were the same colour
it was agreed that the name of the ship would be Titanic
the pension payments were worth less and less every month
her cousin had informed the doctor of his symptoms
the soldier saluted the flag with his rifle by his side
the feast began to get livelier some time later
there were forks in the drawer
the students thought the museum was very boring
the kiln was hot enough to fire the pots
he replied that the songs were quite good
the horn was so loud that they all jumped at the noise
the building had a nest in its roof
the flag was raised to the top of the flagpole
we had to be careful that the ferry was on time
the soldiers thought that helmets would save their lives
the patient bears many injuries this year
a new shopping mall was built last year
we were disappointed that the cookies had not been touched
we were lucky that the hammer was kept in the toolbox
he left school before he had done his exams
his face showed that his team had lost the game

her new skirt was made of denim
the traffic on the highway was very heavy
there was beer and cider on the kitchen shelf
the competition ended as a draw
soccer is mostly played in the summer
her mother was making a cake
the new owners of the house painted it pink
the gate to the church was quite rusty and difficult to open
the king was making many enemies
the care given by the nurses on the ward was very professional
the goal was scored by a defenseman
gin was not a drink that her old man liked
the scouts and the guides always went on long hikes in the summer
snow is unusual in the summer in most countries
awards are given to good writers at the end of their careers
the guard failed to prevent the escape
the panel were supposed to ignore the height and weight of the contestants
the restaurant was bought by the hotel
the wax from the candle fell on the book
the dessert was put into the oven at the start of the meal
he broke his leg when he fell off the horse
his wig fell on the floor
the student tried to move the desk
the noise was very loud and difficult to ignore
the boy was able to conceal his cigarette
he reminded his parents about the game of football
the sketch showed that the road would pass the school
the whole sky was full of birds
the bathroom was decorated by the family to help them to sell the house
there was lettuce and cucumber in the salad
the luggage should be kept in a large warehouse
the boy was able to climb the mountain
the athlete tried to win the marathon
his boss played golf nearly every weekend
the fight in the playground was over a packet of gum
actors normally perform at the theatre
her backpack was full of things that she would need for her camping trip
the singer was well known throughout Europe
there were tools made from gold found at the site
he deserved the respect of his colleagues
the recipe for the cake was easy to follow
she was sitting on the sofa in her bedroom
the garlic and the herbs were added to the fried onion
the statue had some paint on it
she hurt her ankle while she was cycling to the village

the wife of the priest helped out the elderly
the children were hoping to play some hockey and rugby at their school
the fireman climbed down into the bottom of the tunnel
the car drove over the cliff
spiders are often found in the tub
they walked from the cottage down the path to the edge of the forest
he always read a book before going to bed
the pattern on the rug was quite complex
he surprised his parents by his lack of concern
the burglar came up over the wall of the palace
the cattle were kept in the barn
the fog in the valley was quite thick
the computer was sent back after the first month
he added milk and sugar to his coffee
the cake and the biscuits had the same flavour
she laughed at the joke about the dog
the shop was closed when she arrived there
the housewife was able to carry the bags of food
the church was destroyed by the blaze
the money for the science library was increased when the university was modernized
she cleaned the wardrobe after she emptied it
the goat was as greedy as the family had expected
his uncle had some sheep that lived out in his garden
her children saw a snake at the picnic
the gifts sold to the tourists in the shop were quite cheap
the student wrote many essays that year
he met his father while he was walking to the shops
they told the truth about the fight to the teacher
his briefcase was brown and was made of leather
it was a sunny day and the children were going to the park
the camel was kept in a cage at the zoo
the police returned to the museum
the man read the newspaper at lunchtime
he was sitting at his desk in his office
the couple had been together for three years
the child left all of his lunch at home
the soup was kept in a carton in the fridge
some ice was added to the whisky

Appendix B: Coordinates for main effect of task

Contrast	MNI Coordinates (mm) x	y	z	F	Voxels in Cluster	Location	Simple Effect
Main Effect Of Task	44	-68	12	309.09	9756	R middle occipital gyrus	MOT
	-14	-64	56	308.16		L superior parietal gyrus	MOT
	-22	-72	38	261.73		L superior parietal gyrus	MOT
	-38	-14	2	202.90	2901	L superior temporal sulcus	Speech
	-36	-22	6	166.65		L superior temporal sulcus	Speech
	-40	-16	18	153.04		L superior temporal sulcus	Speech
	32	-4	52	191.14	1255	R middle frontal gyrus	MOT
	20	2	52	111.02		R superior frontal gyrus	MOT
	14	-4	70	88.15		R superior frontal gyrus	MOT
	-22	-2	62	180.05	875	L superior frontal gyrus	MOT
	-24	-4	52	143.98		L middle frontal gyrus	MOT
	-16	-8	66	137.52		L superior frontal gyrus	MOT
	-34	-22	46	175.09	176	L central sulcus	Speech
	-50	-14	46	53.20		L central sulcus	Speech
	58	0	4	160.84	3294	R superior temporal gyrus	Speech
	46	-10	4	160.63		R superior temporal gyrus	Speech
	42	-14	10	157.57		R insular cortex	Speech
	18	-28	14	146.71	219	R caudate	MOT
	18	-36	-12	131.95	1662	R parahippocampal gyrus	Speech
	20	-46	0	130.40		R lingual gyrus	Speech
	12	-18	28	126.63		R cingulate gyrus	Speech
	-18	-30	12	104.76	161	L caudate	MOT
	-14	-24	18	77.92		L caudate	MOT
	-24	-36	-44	104.14	228	cerebellum	MOT
	-16	-50	-46	81.04		cerebellum	MOT
	48	6	34	102.24	159	R precentral gyrus	MOT

52	10	26	93.01		R precentral gyrus	MOT
-14	-42	-6	94.49	665	L parahippocampal gyrus	Speech
4	-72	36	80.82		R precuneus	Speech
-2	-78	38	70.85		L precuneus	Speech
-52	38	6	91.27	315	L inferior frontal gyrus	Speech
-42	32	6	72.52		L inferior frontal gyrus	Speech
-48	36	-2	70.40		L inferior frontal gyrus	Speech
-26	-58	-32	85.20	85	cerebellum	MOT
-36	-68	-28	46.25		cerebellum	MOT
-22	-28	66	79.27	41	L postcentral gyrus	Speech
-4	-12	52	75.69	150	L superior frontal gyrus	Speech
26	-38	-44	68.98	111	cerebellum	MOT
16	-48	-46	64.21		cerebellum	MOT
20	-40	-54	50.64		cerebellum	MOT
10	-48	34	68.66	48	R precuneus	Speech
-6	-74	-40	67.68	37	cerebellum	MOT
54	40	4	67.24	32	R inferior frontal gyrus	Speech
8	44	56	65.77	7	R superior frontal gyrus	Speech
-22	-18	64	55.19	14	L precentral gyrus	Speech
4	-36	36	54.08	20	R cingulate gyrus	Speech
-36	-44	-54	54.02	22	cerebellum	MOT
-50	6	34	53.18	10	L precentral gyrus	MOT
-48	20	24	52.65	9	L inferior frontal gyrus	Speech
-12	-52	34	52.25	6	L precuneus	Speech
-14	-28	26	51.96	18	L cingulate gyrus	Speech
-8	-74	-20	50.65	21	cerebellum	MOT
-34	4	12	48.93	5	L inferior frontal gyrus	Speech
-34	-28	-10	48.83	5	L hippocampus	Speech
-30	-16	-20	48.08	4	L hippocampus	Speech
4	36	6	47.11	3	R cingulate gyrus	Speech
14	-66	22	47.00	7	R superior parietal gyrus	Speech

34	18	8	46.06	2	R insular cortex	MOT

Appendix C: Coordinates for simple effect of speech type

	MNI Coordinates (mm)						
Contrast	x	y	z	F	Voxels in Cluster	Location	Simple Effect
Simple Effect of Speech	-50	-18	6	38.98	875	L superior temporal gyrus	CL > NV12
	-60	-10	-4	33.15		L superior temporal gyrus	Intell
	-50	12	-16	20.63		L superior temporal gyrus	Intell
	36	24	-8	36.62	633	R insular cortex	Deg
	36	24	8	20.38		R inferior frontal gyrus	Deg
	-52	-40	2	33.59	207	L middle temporal gyrus	Intell
	-26	22	-8	32.01	318	L insular cortex	Deg
	-34	20	4	18.67		L insular cortex	Deg
	-44	-16	-4	30.41	87	L superior temporal gyrus	Deg
	56	-10	-12	30.32	509	R middle temporal gyrus	Intell
	50	-14	6	21.04		R superior temporal gyrus	Intell
	58	-2	2	20.92		R superior temporal gyrus	Intell
	4	30	52	23.20	414	R superior frontal gyrus	Other
	10	18	62	21.33		R superior frontal gyrus	Other
	8	22	44	17.21		R superior frontal gyrus	Other
	-4	-22	30	20.42	119	L cingulate gyrus	Other
	-42	-12	-30	20.31	36	L inferior temporal gyrus	Intell
	56	-56	20	19.13	112	R angular gyrus	Other
	-28	-8	-20	18.27	20	L hippocampus	Intell
	-34	-36	-18	18.12	202	L fusiform gyrus	Intell
	-24	-38	-12	13.94		L parahippocampal gyrus	Intell
	50	14	22	18.09	74	R inferior frontal gyrus	Other
	68	-28	14	17.92	76	R superior temporal gyrus	Other
	32	-32	-16	17.29	97	R parahippocampal gyrus	Intell
	26	-38	-12	15.83		R parahippocampal gyrus	Intell

44	-70	28	17.19	86	R middle occipital gyrus	Intell
10	-52	12	17.06	28	R precuneus	Intell
50	48	-6	16.86	37	R inferior frontal gyrus	Other
-12	20	66	16.26	18	L superior frontal gyrus	Other
50	18	-24	14.63	10	R superior temporal gyrus	Intell
48	40	26	14.55	18	R middle frontal gyrus	Other
58	12	-14	13.92	4	R superior temporal gyrus	Intell
-30	-22	-16	13.24	3	L hippocampus	Intell
-10	20	36	12.58	2	L superior frontal gyrus	Other
-58	-60	18	12.52	3	L angular gyrus	Intell

Appendix D: Coordinates for tracking-load dependent activity

Contrast	MNI Coordinates (mm) x	y	z	F	Voxels in Cluster	Location	Simple Effect
Tracking Load Correlation	40	-32	46	202.76	3455	R postcentral gyrus	Pos
	26	-88	18	179.18		R superior occipital gyrus	Pos
	30	-82	12	176.80		R middle occipital gyrus	Pos
	-26	-92	12	104.52	1430	L middle occipital gyrus	Pos
	-20	-80	-12	98.46		L inferior occipital gyrus	Pos
	-22	-68	38	95.68		L superior parietal gyrus	Pos
	-42	-20	22	67.67	32	L supramarginal gyrus	Neg
	26	-76	-6	65.50	179	R inferior occipital gyrus	Pos
	12	-82	-12	56.69		R lingual gyrus	Pos
	20	-80	-10	53.92		R inferior occipital gyrus	Pos
	48	-62	26	63.49	77	R angular gyrus	Neg
	60	-56	26	47.29		R angular gyrus	Neg
	8	20	42	60.61	9	R superior frontal gyrus	Pos
	54	12	28	59.10	19	R precentral gyrus	Pos
	24	-30	8	58.42	7	Pulvinar	Pos
	44	44	28	57.36	41	R middle frontal gyrus	Pos
	-44	-60	32	50.20	56	L angular gyrus	Neg
	-6	-86	-10	50.12	6	L lingual gyrus	Pos
	26	-4	56	48.83	50	R superior frontal gyrus	Pos
	32	-62	-20	48.75	6	cerebellum	Pos
	-14	-66	54	48.68	12	L superior parietal gyrus	Pos
	10	-50	30	48.39	10	R cingulate gyrus	Neg
	-38	-42	40	47.58	15	L superior parietal gyrus	Pos
	-26	-16	-14	46.86	1	L hippocampus	Neg
	26	8	56	46.52	7	R superior frontal gyrus	Pos

40	-58	-16	46.38	2	R fusiform gyrus	Pos
10	-58	2	45.17	1	R lingual gyrus	Neg
-26	-70	-52	44.73	2	cerebellum	Pos
-4	-64	18	44.70	1	L precuneus	Neg
10	-58	16	43.75	2	R precuneus	Neg
-4	-74	32	43.57	1	L precuneus	Neg

Appendix E: Speech Type × Task interaction

	MNI Coordinates (mm)						
Contrast	x	y	z	F	Voxels in Cluster	Location	Mask
Speech x Task Interaction	36	24	-4	15.68	434	R Anterior Insula	Speech
	-32	26	-8	14.54	457	L Anterior Insula	Speech
	-34	22	2	13.26		L Anterior Insula	Speech
	-42	22	-2	12.02*		L Inferior Frontal Gyrus	Speech

Note: asterisk indicates $.05 < p < .10$

Appendix F: Speech Type × Task × Load interaction

	MNI Coordinates (mm)						
Contrast	x	y	z	F	Voxels in Cluster	Location	Mask
Speech x Task x Load Interaction	52	4	-12	22.52	655	R anterior superior temporal gyrus	Speech
	-54	4	-12	15.54	548	L anterior superior temporal gyrus	Speech
	48	36	-8	12.05	596	R inferior frontal gyrus	IFG
Speech x MOT Load Interaction	50	32	-6	17.59	1220	R inferior frontal gyrus	IFG
	44	46	-2	13.92		R inferior frontal gyrus	IFG
	-52	16	16	10.58 *	1286	L inferior fontal gyrus	IFG

Note: asterisk indicates $.05 < p < .10$

Harrison Ritz

EDUCATION

Sept. 2014 – (current)	**Master of Science** - Psychology University of Western Ontario in London, ON, CA Specialization: Cognitive and Behavioral Neuroscience Supervisor: Dr. Ingrid Johnsrude Thesis: *The effects of concurrent cognitive load on the processing of clear and degraded speech.*
Sept. 2010 – Apr. 2014	**Bachelor of Science (Honours) with Distinction** - Psychology Queen's University in Kingston, ON, CA GPA (4.3 scale): 3.64 overall, 3.79 in final two years Honors Thesis Supervisor: Dr. Ingrid Johnsrude Thesis: *Attention enhances phase-locking in the brainstem frequency-following response*

CONFERENCE PRESENTATIONS

Ritz, H., Wild, C., Johnsrude, I. J. (2016). *The effects of concurrent cognitive load on the processing of clear and degraded speech.* Organization for Human Brain Mapping 2016. Geneva, CH. Poster Presentation.

Ritz, H., Arbuckle, S., Wild, C., Johnsrude, I. (2015). *Enhanced recognition memory for acoustically degraded sentences.* 39th MidWinter meeting of the Association for Research in Otolaryngology. San Diego, US. Podium presentation; Dr. Johnsrude to give presentation.

Ritz, H., Arbuckle, S., Wild, C., Johnsrude, I. (2015). *Enhanced recognition memory for acoustically degraded sentences.* The Brain and Mind Institute Symposium. London, CA. Poster presentation.

Ritz, H. & Johnsrude, I. (2014). *Attention enhances phase-locking in the frequency following response.* 24th Annual Meeting of the Canadian Society of Brain, Behaviour, and Cognitive Science. Toronto, CA. Poster presentation.

Ritz, H. & Johnsrude, I. (2014). *Attention enhances phase-locking in the frequency following response.* McMaster University NeuroXchange Conference. Hamilton, CA. Poster presentation.

ACCOLADES AND SCHOLARSHIPS

Apr. 2014	**Certificate of Academic Excellence**, Canadian Psychological Association Awarded for the top psychology honours thesis of the year
2011, 2013, 2014	**Dean's Honour List**, Queen's University
Apr. 2011 – Aug. 2011	**Summer Work Experience Program** ($2500)

	Competitive; funded RA position with Dr. Monica Castelhano
Sept. 2010 – Apr. 2014	**Foresters Competitive Scholarship** ($8000)
Sept. 2010	**Queen’s University Excellence Scholarship** ($2000)

RESEARCH ASSISTANTSHIPS

Jan. 2013 – Apr. 2014	**CoNCH Lab**, Queen’s University Supervisor: Dr. Ingrid Johnsrude I designed and collected data for experiments on degraded speech comprehension, and completed my thesis on attentional modulation of brainstem frequency encoding using electroencephalography. Working with Dr. Johnsrude has enriched my critical thinking and methodological rigor.
Sept. 2011 – Dec. 2012	**Queen’s Visual Cognition Lab**, Queen’s University Supervisor: Dr. Monica Castelhano I designed, coded and collected data for several studies on visual search, scene recognition, and visual masking. This position helped to develop my technical skills, and taught me autonomy and responsibility.
July 2010 – Aug. 2010	**Juravinski Cancer Centre**, Juravinski Hospital, Hamilton, ON Supervisor: Dr. Jehonathan Pinthus I was responsible for maintaining and performing chemical assays on several carcinogenic renal cell lines. In this position I learned the value of organization and documentation when following complex experimental protocols.

TEACHING AND MENTORSHIP

Sept. 2015 – Apr. 2016	**PSYC 4851E: Honors Thesis**, Western University Co-supervised honors thesis student
Sept. 2015 – Dec. 2015	**PSYC 3800F: Statistics using Computers**, Western University, TA
Jan. 2015 – Apr. 2015	**PSYC 2115B: Sensation and Perception**, Western University, TA
Sept. 2014 – Dec. 2014	**PSYC 2115A: Sensation and Perception**, Western University, TA
Sept. 2012 – Apr. 2013	**PSYC 100: Introduction to Psychology**, Queen’s University, TA

SERVICE

Sept. 2015 - Apr. 2016	**Psychology Graduate Students Association**, Western University Executive: I am responsible for communication, advocacy, and event planning for psychology and neuroscience graduate students.

May 2015 - Apr. 2016	**Psychology Colloquium Committee**, Western University Committee Member: I participated in selecting the colloquium speakers for the 2015/2016 season, and organized the logistics for a talk by Dr. Adam Anderson in November, 2015.
Sept. 2013 – Apr. 2014	**Psychology Departmental Student Council**, Queen's University 4th year representative: I attended weekly student council meetings and monthly undergraduate departmental meetings, helped organize and run social events, and administered instructor evaluations.
Sept. 2012 – Apr. 2014	**Queen's Cinema Appreciation Society**, Queen's University Founder and President: I curated and organized screenings, and led discussions on a weekly film from a diverse range of genres, languages, and time-periods.
Sept. 2011 – Apr. 2014	**Campus Observation Room**, Queen's University Volunteer: I assessed heavily intoxicated students against our admittance criteria, and monitored these students' ongoing condition. I also assisted with alcohol safety workshops and promotional campaigns.