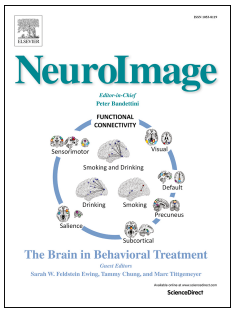
Accepted Manuscript



Anterior paracingulate and cingulate cortex mediates the effects of cognitive load on speech sound discrimination

Silvia P. Gennari, Rebecca E. Millman, Mark Hymers, Sven L. Mattys

PII: S1053-8119(18)30542-1

DOI: 10.1016/j.neuroimage.2018.06.035

Reference: YNIMG 15041

To appear in: *NeuroImage*

Received Date: 10 October 2017

Revised Date: 7 June 2018

Accepted Date: 10 June 2018

Please cite this article as: Gennari, S.P., Millman, R.E., Hymers, M., Mattys, S.L., Anterior paracingulate and cingulate cortex mediates the effects of cognitive load on speech sound discrimination, *NeuroImage* (2018), doi: 10.1016/j.neuroimage.2018.06.035.

Anterior paracingulate and cingulate cortex mediates the effects of cognitive load on speech sound discrimination

Silvia P. Gennari[1], Rebecca E. Millman[2,3], Mark Hymers[1], and Sven L. Mattys[1]

[1]*Department of Psychology, University of York, YO10 5DD, UK*

[2]*Manchester Centre for Audiology and Deafness, School of Health Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester, M13 9PL, UK*

[3] *NIHR Manchester Biomedical Research Centre, Central Manchester University Hospitals Foundation Trust, Manchester Academic Health Science Centre, M13 9WL, UK.*

Corresponding author

Sven Mattys sven.mattys@york.ac.uk

Department of Psychology, University of York, YO10 5DD, UK

Abstract

Perceiving speech while performing another task is a common challenge in everyday life. How the brain controls resource allocation during speech perception remains poorly understood. Using functional magnetic resonance imaging (fMRI), we investigated the effect of cognitive load on speech perception by examining brain responses of participants performing a phoneme discrimination task and a visual working memory task simultaneously. The visual task involved holding either a single meaningless image in working memory (low cognitive load) or four different images (high cognitive load). Performing the speech task under high load, compared to low load, resulted in decreased activity in pSTG/pMTG and increased activity in visual occipital cortex and two regions known to contribute to visual attention regulation—the superior parietal lobule (SPL) and the paracingulate and anterior cingulate gyrus (PaCG, ACG). Critically, activity in PaCG/ACG was correlated with performance in the visual task and with activity in pSTG/pMTG: Increased activity in PaCG/ACG was observed for individuals with poorer visual performance and with decreased activity in pSTG/pMTG. Moreover, activity in a pSTG/pMTG seed region showed psychophysiological interactions with areas of the PaCG/ACG, with stronger interaction in the high-load than the low-load condition. These findings show that the acoustic analysis of speech is affected by the demands of a concurrent visual task and that the PaCG/ACG plays a role in allocating cognitive resources to concurrent auditory and visual information.

Keywords: speech perception, cognitive load, visual memory load, divided attention, anterior cingulate gyrus, fMRI

Introduction

Much research on speech perception has been devoted to characterizing the functional network and processing streams involved in perceiving and comprehending spoken language (e.g., Hickok & Poeppel, 2007; Friederici, 2011; Price, 2012). For example, dorsal and ventral processing streams have been identified to account for the mapping of speech sounds to meaning and associated motor gestures. However, speech perception often occurs in the context of competing visual information, whereby the brain must strategically allocate processing resources to both modalities simultaneously. How attentional control under divided attention interfaces with speech processing neural networks is not well understood. In particular, it is unclear whether (1) Brain areas associated with the early analysis of speech perception are affected by a secondary, non-auditory task and (2) Language-related pre-frontal regions such as the left inferior frontal gyrus and other executive control regions such as dorsolateral pre-frontal cortex (e.g., Miller and Cohen, 2001) play a role in monitoring attention allocation to speech while dual-tasking. The present study addresses these questions.

Processing speech under focused attention and in the absence of distraction is known to involve bilateral temporal brain regions, especially the posterior superior and middle temporal gyri (pSTG and pMTG) and the superior temporal sulcus (STS) (e.g., Scott et al., 2000; Davis & Johnsrude, 2003; Hickok & Poeppel, 2007; Obleser et al., 2007; Price, 2012). However, when visual stimuli are presented simultaneously with auditory stimuli, selective attention to one modality results in increased activity in the sensory cortex corresponding to that modality and decreased activity in the sensory cortex corresponding to the unattended modality (Downar et al., 2001; Johnson & Zatorre, 2005, 2006; Sabri et al., 2008; Molloy et al., 2015). Decreased activity corresponding to the unattended modality is also found subcortically (Rinne et al., 2008; Sörqvist et al., 2012). Thus, selective attention is implemented both cortically and sub-cortically, with limited involvement of frontal executive regions.

Compared to selective attention, divided attention modulates sensory cortex activity to a lesser degree, but it enlists a variety of frontal regions (Loose et al., 2003; Shomstein & Yantis, 2004; Johnson, 2006; Johnson et al., 2007). It has been argued that, under divided attention, recruitment of left prefrontal regions compensates for decreased sensory activity, allowing similar levels of performance to be maintained (Johnson & Zatorre, 2006). However, the mechanisms underlying resource allocation during divided attention are unclear, as the pattern of results is greatly constrained by the nature of the component tasks (Salo et al., 2015). Furthermore, contrasting focused and divided attention as a way of understanding resource allocation during divided attention can be problematic because brain activity differences could result as much from a voluntary change in strategy under dual-tasking as from automatic resource trade-offs (Navon, 1984; Navon & Gopher, 1979).

Resource allocation under divided attention can be more effectively investigated by keeping the main task constant (speech processing) while manipulating the degree of cognitive load generated by the concurrent task (e.g., the degree of complexity of a visual working memory task). Behavioral evidence shows that speech processing accuracy decreases as a function of the difficulty of a concurrent task (Brungart et al., 2013; Francis, 2010; Mattys & Wiget, 2011; Mattys et al., 2014) and that the locus of this decrement is perceptual rather than decisional (Macdonald & Lavie, 2011; Mattys & Palmer, 2015; Raveh & Lavie, 2015). This leads us to ask whether allocation of resources during speech perception under cognitive load is implemented primarily via a change of activity in the sensory cortices or whether it also involves frontal regions to handle cross-modal attentional allocation. In the first case, we would expect the complexity of the concurrent task to affect primarily the brain areas associated with the acoustic analysis of speech (pSTG, pMTG). In the second case, we would expect it to also involve frontal regions.

The present study used functional magnetic resonance imaging (fMRI) and a cross-modal divided attention paradigm (speech perception task + visual working memory task) to investigate

the neural networks involved in regulating speech perception as a function of the difficulty of a concurrent cognitive load. In keeping with Hickok and Poeppel (2007), we use the term "speech perception" to refer to the sublexical processes involved in the acoustic-phonetic analysis of the speech signal, as opposed to "speech recognition," which encompasses processes involved in mapping sounds to lexical representations and meaning. The speech task we used was AX phoneme discrimination performed on two syllables played one after the other (/gɪ-kɪ/). We chose this specific task because it has a long history of tapping into the building blocks of speech perception (e.g., Hutchison, Blumstein, & Myers, 2008), independent of word recognition and sentence comprehension (e.g., Baker, Blumsteim, & Goodglass, 1981).

The concurrent visual task consisted of the rapid serial visual presentation of four un-nameable and meaningless images during the playback of the two spoken syllables. The four images were either the same (low load) or different (high load). At the end of each trial, participants reported whether the two spoken syllables were the same or different (the speech task) and, then, whether or not a probe image presented at the end of the trial was among the images displayed during the trial (the visual task). The visual task allowed us to load visual working memory at two contrasted degrees of magnitude (4 items vs. 1 item) while maintaining divided attention in both conditions. The fact that the images were un-nameable ensured that they could not be rehearsed sub-vocally, and hence, that their representations in working memory did not interfere with the representations of the spoken syllables. BOLD responses were only collected for the period of time corresponding to the presentation of the spoken syllables so that conflict or uncertainty at the time of responding was not recorded. This allowed us to specifically target working memory load and allocation of attention during stimulus processing, independently of response conflict typically associated with frontal lobe activity. Since the visual working memory task was designed to modulate cognitive demands during the divided-attention task, we refer to our conditions as high vs. low cognitive load conditions.

We predicted that, if cognitive load modulates early analysis of the speech signal, as might

be expected from selective attention research (e.g., Molloy et al., 2015; Sabri et al., 2008), activity in pSTG/pMTG should decrease when cognitive load increases. Alternatively, a lack of activity change in these regions in response to load would suggest that the reduction in perceptual sensitivity observed in behavioral studies is not implemented at the acoustic-phonetic level of sound processing but at a higher level of processing. Furthermore, if the interaction between cognitive load and early analysis of the speech signal is modulated by executive or control regions, we expect left frontal areas to be engaged more strongly in the high-load condition than in the low-load condition and for that effect to be functionally linked to the pSTG/pMTG. Finally, we reasoned that if performance in the visual task depends in part on the ability to deal with divided attention in demanding conditions, individual performance should be predicted by the activity level in executive control regions.

Methods

Participants

Eighteen native English-speaking, right-handed adult volunteers (mean age 23.6 years; range 18 to 34; 9 females) with no reported hearing or visual deficits gave informed consent to participate in the study. The project was approved by the Research Governance Committee, York Neuroimaging Centre, University of York, and conformed to the guidelines given in the Declaration of Helsinki.

Materials

The auditory stimuli were two spoken syllables. Both had a consonant-vowel structure and shared the same vowel (/ɪ/). The initial consonant was a bilabial plosive with a voice onset time of 23 msec vs. 48 msec. The two syllables sounded like a somewhat ambiguous version of /gɪ/ (as in “gill”) and /kɪ/ (as in “kill”), respectively. These two syllables, which were used in Mattys and Wiget (2011), were chosen because pilot data indicated that discrimination between them in

simulated scanner background noise at 0 dB SNR was around 75%. This performance level was judged to be adequately different from both chance and ceiling. The syllables were created by recording a natural syllable and reducing or increasing voice onset time through manual sound editing (see Mattys & Wiget, 2011, for full details). Syllable duration was 109 msec for /gɪ/ and 134 msec for /kɪ/. The average fundamental frequency of the vowel was 224 Hz.

The visual images used for the cognitive load task were 30 different characters (or character components) drawn from the Bengali, Tamil, and Gujarati alphabets (see examples in Figure 1). The size of each image was approximately 3x3 cm. Images were displayed in white against a black background.

fMRI data acquisition

Data were acquired using Interleaved Silent Steady-State Imaging (ISSS) (Schwarzbauer et al., 2006). Details of the approach to ISSS data acquisition and data analysis are described in Hymers et al. (2015). ISSS allows for the acquisition of multiple temporal volumes after a quiet period without the necessity of modelling T1 saturation effects. It has been shown to be more sensitive than traditional sparse imaging when performing auditory fMRI experiments (Mueller et al., 2011).

During the acquisition periods, whole-head fMRI data (GE-EPI, TR = 2 s, TE = Minimum Full, flip angle = 90º) were collected using a GE Signa HDx 3T system (General Electric, Waukesha, WI, USA). A 64 x 64 pixel matrix with a field of view (FOV) of 19.2 cm was used, giving an in-plane resolution of 3 mm x 3 mm. Thirty-eight interleaved slices were collected with a slice thickness of 3 mm. A total of 148 3D volumes of data were acquired for each run, and each participant completed four runs. The 8 initial non-acquisition volumes during the first pre-presentation accounted for initial T1 saturation effects. High-resolution T1-weighted anatomical scans (3D Fast Spoiled Gradient-Recalled Echo) were obtained for each participant (TR = 7.8 msec, TE = Minimum Full, TI = 450 msec, flip angle = 20º, voxel size = 1.13 mm x 1.13 mm, matrix size

= 256 x 256 x 176, giving a FOV of 290 mm slice; thickness = 1 mm).

Design and procedure

The experiment was divided into four consecutive fMRI runs (or scanning sessions) of 9 min 40 sec each. Each run contained a total of 32 trials and was further sub-divided into four mini-blocks containing eight trials each. Mini-blocks alternated between low-load and high-load. Each mini-block was followed by a rest period of 16 s to let the hemodynamic response decrease to baseline. The order of low-load and high-load mini-blocks was counterbalanced across participants. Since each trial required performance on two tasks with two possible alternative responses each (*same*/*different* sound, *present/absent* image), the proportion of correct response combinations was the same in each run (*same-present, same-absent, different-present, different/absent*). A given combination of correct responses never occurred in two or more consecutive trials and a given task response (e.g., *same* or *different*) never occurred in more than three consecutive trials.

Each trial lasted 16 sec, as illustrated in Figure 1. At the start of each trial, there was a rest period of 6 sec during which a fixation cross was displayed. After the rest period, the auditory and visual stimuli were presented simultaneously for 2 sec. During this period, no fMRI data were acquired. Stimulus presentation was followed by the acquisition of four fMRI volumes (TR = 2 s), i.e., 8 sec of data acquisition during each trial. Visual stimuli were viewed through an angled mirror placed on the head coil. The auditory stimuli were played binaurally at 98 dB SPL over noise-attenuating fMRI-compatible headphones (MR Confon, MR Confon GmBH). This presentation level does not take into account the attenuation provided by earplugs worn throughout the experiment (approximately 15 dB SPL). The sound level of the scanner noise during the quiet period was 81 dB SPL and, during the acquisition period, the background noise level was 98 dB SPL.

The auditory and visual stimuli were presented using Neurobehavioural Systems Presentation version 13.1. The visual stimuli consisted of four consecutive images, each displayed

for 350 msec, with approximately 30 msec of black screen between images. The first image in the series was presented 150 msec after trial onset. In the low-load condition, the four images were the same. In the high-load condition, they were different from each other. During the presentation of the images, two spoken syllables were played, with a 500-msec inter-stimulus interval. The first syllable was presented 500 msec after trial onset. The syllables were either the same (/gɪ/-/gɪ/ or /kɪ/-/kɪ/) or different (/gɪ/-/kɪ/ or /kɪ/-/gɪ/). The auditory and visual stimuli were aligned on their mid-duration, that is, the visual presentation had a short lead and lag time before and after the auditory stimuli. Twelve seconds into the trial (see Figure 1), an image representing headphones and a question mark appeared on the screen, which prompted participants to report whether they thought the two spoken syllables were the same or different. Participants provided responses using a button box placed under their right hand, with the index finger for *different* and the middle finger for *same*. Two seconds later, an image appeared on the screen together with a question mark, prompting participants to indicate whether the image was among the images displayed earlier in the trial. Participants provided responses to the visual task using a button box placed under their left hand, with the index finger for *absent* and the middle finger for *present*. Participants were instructed to perform both tasks as well as they could, i.e., one task was not prioritized over the other.

fMRI Analysis

The fMRI data were analysed using Feat-5.98, part of FSL 4.1 (FMRIB's Software Library, http://www.fmrib.ox.ac.uk/fsl), as well as custom scripts which implemented filtering of the temporally non-contiguous data (Hymers et al., 2015). Functional data were first co-registered to a high-resolution T1-weighted anatomical image for individual participants and then non-linearly co-registered onto a standard MNI brain (ICBM152). A separate first-level analysis was carried out for each run and each participant. This allows the application of motion correction for each run separately, as head motion between scanning sessions may occur, even for consecutive runs. Motion-correction was implemented using MCFLIRT (Jenkinson et al., 2002) and brain extraction

was performed using BET (Smith, 2002). The motion correction parameters were entered as regressors of no interest in the general linear model. Regressors of no interest associated with cerebrospinal fluid and white matter were obtained, as described in Barton et al. (2015), and included in the model of all analyses discussed below. Spatial smoothing was performed on the EPI data using a Gaussian kernel full-width half-maximum of 6 mm. Linear and quadratic trends were removed per-voxel using an in-house tool, which took into account the times at which data were acquired (Hymers et al., 2015).

There were four experimental explanatory variables: (1) auditory same with low cognitive load, (2) auditory different with low cognitive load, (3) auditory same with high cognitive load, and (4) auditory different with high cognitive load. Due to the nature of the ISSS acquisition sequence, the non-contiguous temporal structure of the acquired EPI data had to be taken into account when performing analyses. To do so, we used a version of the analysis pipeline described in Peelle (2014). The time series of events was initially constructed to span the entire length of the experiment regardless of data acquisition. Each event in the series consisted of the 2-sec period from the onset of the stimulus and was convolved with a double gamma haemodynamic response function along with its temporal derivative (Friston et al., 1998). Regressors only accounted for activity during stimulus processing and did not include later button presses. Taking into account the lag of the hemodynamic response, the four acquisition volumes in Figure 1 only captured the stimulus presentation activity. The time series was then re-sampled at the time of fMRI data acquisition using a local modification of the standard FSL analysis routines. The appropriate regressor heights were recalculated for the explanatory variables and their contrasts to take into account the temporally reduced time series. The resulting design matrix was used with FMRIB's Improved Linear Model (FILM) in order to estimate beta values. Contrasts of parameter estimates for high- vs. load-load comparisons were calculated by pooling together the same and different response trials.

Parameter estimates were then carried through to a second-level fixed-effects analysis in

which the four runs of each participant were combined. Finally, data from individual participants were entered into a third-level mixed-effects analysis using FLAME (FMRIB's Local Analysis of Mixed Effects, Beckmann et al., 2003; Woolrich et al., 2004). Three regressors were used in the third level analysis: One for the mean effect of cognitive load and two covariates corresponding to the demeaned per-participant performance on the auditory and visual tasks to account for potential variability in BOLD responses caused by performance. The main effect of cognitive load was orthogonalized with respect to these behavioural regressors, thus removing the influence of performance from the main manipulation of interest. The primary contrast of interest was the mean effect of high-load > low-load. This contrast was evaluated in cluster-corrected whole-brain analyses, where *Z* (Gaussianized T/F) statistic images were thresholded using clusters determined by $Z > 3.1$ and a cluster significance threshold of $p = .05$ (Worsley, 2001).

Region of interest analysis. To assess whether any frontal activity played a role in the task, as hypothesized, we restricted our analyses to a large region of interest (ROI), namely, the left frontal lobe (Johnson & Zatorre, 2006), as defined by the MNI atlas (a mask of this region was entered at the group-level analyses). Within this ROI, *Z* (Gaussianized T/F) statistic images were thresholded using clusters determined by $Z > 3.1$ and a cluster significance threshold of $p = .05$.

Psychophysiological interaction analysis (PPI)

This analysis was conducted to establish the pattern of functional connectivity of auditory regions as a function of cognitive load levels. To define the seed region, we selected the voxel in the posterior temporal lobe that responded most strongly to the cognitive load manipulation (see peak voxel in Table 1) and drew a 6 mm sphere around it. In the PPI model, the physiological regressor corresponded to the first eigenvariate of the sphere's voxels after removing the variance accounted for by regressors of no interest (cerebrospinal fluid and white matter, as above). The psychological regressor was the high-load > low-load contrast (Friston et al., 1997). The psychological regressor was convolved with the gamma function and temporal derivatives with temporal filtering added in

order to simulate the hemodynamic response function. The PPI regressor was the interaction term between the zero-centered psychological regressor and the demeaned physiological regressor. As in the analyses above, behavioral regressors were included in the model and were orthogonal to the main effect of interest. Note that FSL differs from SPM in convolving the psychological regressor, which is treated as a nuisance variable in the interaction, rather than deconvolving the physiological regressor. Although this has been suggested to lead to loss of statistical power for event-related designs, the results that do survive are nevertheless meaningful (Gitelman et al., 2003). To evaluate group statistics across the whole brain, higher-level analyses were carried out using a cluster-thresholding procedure at $Z > 2.0$ and $p < .05$, because we expected relatively low statistical power in the present design.

Correlations

To establish correlations between brain activity and individual differences in behavioral performance, we used FSL utilities to extract the mean percentage signal change for each participant and contrast in those significant clusters resulting from the whole-brain group analysis. To obtain larger clusters that could capture individual variability when transforming from MNI space to individual space, we used a cluster correction threshold of $Z = 2.3$, $p = .05$. Specifically, a cluster within the frontal lobe that responded more strongly to the high-load than the low-load condition was used as a mask to extract the mean percentage signal change for each participant. These values were then correlated with each participant's average behavioral performance. A similar procedure was used for significant clusters found within the superior and middle temporal lobes and the superior parietal lobule.

Results

Behavioral performance

Behavioral performance in the auditory and visual tasks was analyzed using mixed-effects regression models (see Figure 2). Raw data were modeled using a logit function to account for their binomial distribution. For both analyses, the fixed factors were Load (low load, high load) and Type of Auditory Pair (same, different). We did not have specific hypotheses regarding the latter factor. It was simply included to check if the effect of cognitive load was modulated by whether or not the auditory stimuli required a positive or negative answer. All tests included by-participant random intercepts and by-participant random slopes for Type of Auditory Pair. Attempts to model more complex random structures led to convergence failure or near-singular covariance matrices, which can be misleading for interpreting fixed effects (Bates, Kliegl, Vasishth, & Baayen, 2015). The base model included only the random terms. The effects of Load and Type of Auditory Pair were assessed by testing the increase in model fit using the likelihood ratio test when each of these factors was added to the base model.

In the auditory task, Load significantly increased the model fit, $\beta = .34$, SE $\beta = .14$, $\chi^2(1) = 6.02$, $p = .014$, suggesting that auditory discrimination was poorer under high load than low load (87% vs 90%, i.e., 3% cost). This confirms the gradient effect of cognitive load on speech perception, as indicated by Mattys et al. (2011, 2014). Adding Type of Auditory Pair did not improve the fit, $\chi^2(1) = 2.26$, $p = .13$. Neither did adding an interaction term between the two factors, $\chi^2(1) = 1.09$, $p = .30$ (same = 5% cost; different = 2% cost). The generally high syllable discrimination performance compared to our pilot data probably reflects better-than-expected noise attenuation in the scanner.

In the visual task, a main effect of Load, $\beta = 1.10$, SE $\beta = .11$, $\chi^2(1) = 113.51$, $p < .001$, confirmed that the high load was more demanding than the low load (67% vs. 85%, i.e., 18% cost). Adding Type of Auditory Pair did not further improve the fit, $\chi^2(1) = 1.93$, $p = .16$, and neither did the interaction term, $\chi^2(1) = .25$, $p = .62$ (same = 20% cost; different = 16% cost).

fMRI results

Figure 3 shows the results of the whole-brain and ROI analyses. The results for the mean effect of high-load > low-load are shown in warm colors (yellow/red) and those for low-load > high-load are shown in cool colors (light blue/dark blue). Table 1 shows MNI coordinates for maximum voxels in different clusters of activation (anatomical labels were obtained from the Harvard-Oxford cortical structural atlas). In parentheses, Table 1 also shows additional frontal clusters that were significant at a less stringent whole-brain correction level ($Z = 2.3$, $p = .05$).

High-load > low-load contrast. Larger BOLD responses for the high-load than low-load conditions were localized in bilateral occipital complexes (LOC) and fusiform gyri responsible for feature and category-level visual processing (Grill-Spector et al., 2001; Grill-Spector, 2003). This difference reflects greater visual processing in the high-load than low-load conditions. At a more lenient threshold (whole brain cluster corrected, $Z = 2.3$, $p = .05$), significant activity was also found in the right superior parietal lobule (SPL), an area known to be involved in tasks requiring high-visual attention (Wojciulik & Kanwisher, 1999; Hopfinger et al., 2000; Molenberghs et al., 2007).

Analyses conducted within the left frontal lobe ROI revealed no significant differences for the high-load > low-load contrast in lateral prefrontal regions, as previously found. Instead, significant activity for this contrast was found in medial pre-frontal regions at the intersection of the paracingulate gyrus (PaCG) and anterior cingulate gyrus (peak coordinate: x = 2, y = 16, z = 42), with sub-threshold activity also extending into the anterior cingulate gyrus (ACG) (See Table 1). This suggests that control regions were recruited when processing demands increased. The ACG and neighbouring areas are known to be recruited during conflict monitoring and attentional control (Botvinick et al., 2001, 2004; Shenhav et al., 2013; Wei et al., 2013). In our task, the auditory and visual stimuli did not conflict, but the high cognitive load may have required additional resources to handle the divided attention task. In sum, these analyses indicate that, compared to the low-load condition, the high-load condition activated a network of regions involved in visual processing, as well as attentional control regions in the parietal lobule and PaCG/ACG.

Low-load > high-load contrast. Regions responding to this contrast were situated in the

posterior superior and middle temporal lobe and in the lingual gyrus. Critically, left pSTG/MTG, which are associated with acoustic-phonetic speech analysis (Joanisse & Gati, 2003; Myers et al., 2009), showed stronger activation under low load than high load, even though the behavioral results showed better performance in the low-load than high-load conditions. Thus, while challenging speech perception typically leads to increased activity in left posterior temporal regions in single-task paradigms (cf. speech signal distortions in Davis and Johnsrude, 2003), in the present divided attention paradigm, high load led to reduced activity in those regions. This result supports the hypothesis that the activation of speech-related brain regions can be altered by the degree of cognitive load elicited by a concurrent visual task under divided attention, reflecting a modulation of auditory processes by visual task demands.

Greater activity under low than high load was also found in the lingual gyri. The lingual gyri are thought to support visual working memory encoding and maintenance (Machielsen et al., 2000; Vaidya et al., 2002). In our experiment, greater lingual gyrus activity under low load could be explained by the fact that encoding and maintenance of the visual information was only possible when there was only one image to process. The rate of presentation would have made it more difficult to fully encode the visual stimuli in the high-load condition. In that condition, higher-order cortices would have been required because of perceptual interference between the successively presented images.

These analyses show that the high-load condition led to greater involvement of visual areas and heteromodal areas associated with conflict resolution and attentional control (SPL and PaCG/ACG). These two groups of areas are known to be linked via the superior longitudinal fasciculus (Catani et al., 2002). In contrast, left posterior auditory regions responsible for acoustic-phonetic speech analysis (pSTG and pMTG) were more strongly recruited in the low-load condition than in the high-load condition, showing modulation of auditory processing by the demands of a concurrent cognitive load. Since visual and auditory cortices are not known to be anatomically connected to each other through direct white matter tracts, it is likely that this modulation was

mediated by the SPL and ACG.

Correlations between activated clusters. To examine the relationship between the pSTG/pMTG and medial anterior frontal regions across participants, we correlated the mean percentage signal change across all medial frontal voxels (thresholded at $Z = 2.3$) showing modulation by cognitive load (which included cingulate and paracingulate gyri, shown in Figure 3 and Table 1) with the mean percentage signal change of the pSTG/pMTG cluster (see Figure 3). We also examined activity in the SPL cluster to establish whether the PaCG/ACG operated in synchrony with the SPL to deal with the demands of the visual task in the high-load condition. Although ACG/PaCG activity was not correlated with SPL activity ($R^2 = .003$, $p > .90$), it was correlated with activity in pSTG/pMTG (Figure 4A). Specifically, the percentage change for the high-load > low-load contrast in the ACG/PaCG negatively correlated with the percentage change of the same contrast in pSTG/pMTG (*Spearman's* $\rho = -.50$, $p = .04$). Across participants, in the high-load condition relative the low-load condition, greater activity in ACG/PaCG was associated with less activity in pSTG/pMTG. This relationship provides preliminary evidence for an activation trade-off between ACG/PaCG and pSTG/pMTG as a function of cognitive load.

Brain activity and individual behavioral performance. To establish whether brain activity predicted behavioral performance, we regressed the mean percentage signal change in the PaCG/ACG and SPL clusters revealed by the whole brain analyses (cluster corrected, $Z = 2.3$, $p = .05$) against the average percentage of correct responses in the visual task. We restricted our analyses to the visual task because there was too little variance across participants in the auditory task to be predicted by brain activity. We reasoned that if performance on the visual task resulted in part from the attention allocated to the visual stimuli, performance should be predicted by the activity level in executive regions responsible for attentional control such as the ACG and SPL.

We found that the average percentage signal change for the high-load > low-load contrast in the PaCG/ACG cluster predicted performance in the visual task (*Spearman's* $\rho = -.48$, $p = .05$),

whereas the SPL activity showed no relation to performance (*Spearman's* $\rho = .09$, $p = .69$). In particular, as shown in Figure 4B, the larger the activity difference between the high-load and low-load conditions in the ACG/PaCG, the poorer the average performance in the visual task. This suggests that participants who most strongly engaged PaCG/ACG in the high-load condition relative to the low-load condition also experienced more difficulty in the visual task. That is, good performers in the visual task did not engage PaCG/ACG as much as those who performed more poorly. A similar negative relationship, albeit with dorsolateral pre-frontal cortex, has been reported for memory accuracy (Johnson & Zatorre, 2006) and response latency (Rypma & D'Esposito, 1999; Herath et al., 2001).

PPI results

Connectivity between the seed region in pSTG/pMTG and key areas of heteromodal cortex varied as a function of load condition. Activity in the seed region correlated more strongly with a large medial cluster, including PaCG/ACG, in the high-load than the low-load condition (Figure 5 and Table 2). This cluster encompassed portions of the right, middle, and left paracingulate gyrus and left cingulate gyrus, overlapping with the regions reported for the high-load > low-load contrast in Table 1 (36 voxels with centre of gravity at x = -6.86, y = 29.89, z = 21.15). Recall that in the main functional analysis, the pSTG/pMTG region was more strongly activated in the low-load than the high-load condition. Yet, it correlated with PaCG/ACG more strongly in the high-load than the low-load condition. This result thus provides evidence for a modulating role of PaCG/ACG, allocating fewer attentional resources to speech processing regions in the high-load condition.

Discussion

Behavioral research has shown significant effects of cognitive load on auditory processes during speech perception (e.g., MacDonald & Lavie, 2011; Mattys et al., 2014; Mattys & Palmer, 2015; Raveh & Lavie, 2015). More demanding cognitive load results in poorer speech perception. The present study aimed to identify the brain mechanisms underpinning this relationship. In

particular, we sought to determine whether auditory cortical responses to speech are modulated by the complexity of a concurrent visual task, and whether these responses interact with executive regions recruited to compensate for increased cognitive demands in divided attention. We found that performing a visual task with high cognitive demands (holding four images in memory rather than one) resulted in decreased activity in pSTG/pMTG, a region associated with the acoustic-phonetic analysis of speech, and in increased activity in visual processing regions (LOC) and attentional control regions, in particular, PaCG/ACG. Furthermore, activity in PaCG/ACG predicted both performance in the visual task and neural activity in pSTG/pMTG: The greater the activity difference between the high- and low-load conditions in anterior cingulate regions, the less corresponding activity in pSTG/pMTG, and the poorer the performance. Importantly, activity in the pSTG/pMTG seed region showed psychophysiological interactions with portions of the PaCG/ACG, with stronger connectivity in the high-load than low-load condition. Our results suggest that visual cognitive load decreases auditory-related activity via the recruitment of anterior cingulate and paracingulate regions, and this modulation partially accounts for visual behavioral performance in cross-modal divided attention.

The results are consistent with prior reports indicating that activity in auditory areas during a visual memory task decreases more if the visual task is cognitively demanding than if it is less cognitively demanding (Sörqvist et al., 2012; Molloy et al., 2015). Our results extend this finding to divided attention and, critically, they reveal key mechanisms responsible for the flexible deployment of attention to the two concurrent tasks. Even when participants are actively required to engage in an auditory task while performing a visual task, auditory cortical activity trades off with the degree of complexity of the visual task. Our results differ from prior divided-attention studies in that our manipulation did not elicit recruitment of dorsolateral or ventrolateral prefrontal cortex (Iidaka et al., 2000; Loose et al., 2003; Johnson & Zatorre, 2006). However, Loose et al. reported increased ACG activity when comparing divided with selective attention. Our results show that the involvement of the PaCG/ACG extends to managing attentional allocation as a function of

cognitive load within divided attention. One predictor of ACG activity in divided attention is the nature of the task: Stimulus encoding in episodic memory for delayed responses (e.g., after experimental blocks), as in Johnson and Zatorre (2006), does not appear to engage the ACG, whereas stimulus encoding in working memory for immediate trial-by-trial responses does, as was the case in our study. Importantly, the above studies differ from ours in that only ours compared divided attention conditions varying in the cognitive demands required by the non-auditory task. Thus, executive functions dealing with limited processing resources—particularly in the high-load condition—led participants to flexibly engage the PaCG and neighbouring regions during stimulus encoding rather than lateral pre-frontal cortex. These differences across studies reflect the brain's flexible adaptation to graded task demands and suggest that increased cognitive effort in divided attention paradigms require involvement of the PaCG/ACG during stimulus processing.

Several studies have investigated the functions of ACG, especially with respect to the more dorsal portion of the ACG activated in the present study (Picard & Strick, 1996; Beckmann et al., 2009). These functions include conflict monitoring, attentional control—particularly in visual studies dealing with inhibition of distracting stimuli—error detection, updating of predictive models in changing contexts, reward processing, encoding of action relative to costs and gains, and allocation of control based on expected reward (MacDonald et al., 2000; Botvinick, 2004, 2007; Rushworth & Behrens, 2008; O'Reilly et al., 2013; Shenhav, 2013; Ebitz & Hayden, 2016). Our findings in LOC and ACG are consistent with a view of ACG as allocating attention to the visual stimulus: Activity in both these regions increased in the high-load condition relative to the low-load condition, and the relative increase in ACG predicted behavioral performance in the visual task. Since our study did not acquire BOLD data during periods corresponding to response decision, PaCG/ACG activity cannot be attributed to response conflict, error detection, response monitoring, or decision-making behavior based on cost and rewards. Therefore, PaCG/ACG activity must reflect the management of stimulus encoding in order to optimize performance in the two tasks. This control process correlated with the inhibition of auditory-related activity, as evidenced by the

PPI results and correlations between PaCG/ACG and pSTG/pSTG activity changes. Taken together, our results suggest that the PaCG/ACG modulates attention across different modalities, enhancing/inhibiting visual/auditory sensory cortices to meet task demands. Thus, in response to the visual task demands, the PaCG/ACG is recruited in modulating activity across the network to adapt to the current task demands. This is consistent with current proposals suggesting that the ACG is involved in strategic adjustments in cognitive control to adapt to task difficulty (Botvinick et al., 2001, 2004; Shenhav et al., 2013), but further highlights its role in cross-modal attentional control.

The involvement of the PaCG/ACG extends to challenging listening conditions beyond divided attention. For instance, activity in the cingulo-opercular network is shown to increase during word identification in noise (e.g., Eckert et al., 2009; Erb & Obleser, 2013), probably because of the greater need to regulate uncertainty (e.g., Shenhav et al., 2013). Sustained activity in those regions is also necessary for memory maintenance of degraded speech for later retrieval (Vaden et al., 2017). Thus, the ACG and surrounding areas are critical in negotiating a balance between allocating resources to speech-vs-noise segregation, one the one hand, and encoding/maintaining speech in memory, on the other. Although our experiment did not involve speech in noise, the ambiguous syllables we used required effortful listening and probabilistic acoustic judgement, as likely did the stimuli in the studies involving noise. Likewise, discrimination tasks, like the AX task we used, require working memory to maintain the stimuli in an active state for the duration of the trial (Hickok & Poeppel, 2007; Price, 2012). Therefore, enhanced activation of the ACG under high load suggests attempts to regulate perception and working memory processes.

The role of the ACG in cross-modal attention allocation in divided attention paradigms is likely linked to its architectural properties such as its medial location and connectivity to subcallosal structures (e.g., striatum) and to pre-motor cortex (Beckmann et al., 2009). White matter tracts in the corpus callosum also connect the two hemispheres and carry information from sensory inputs to frontal regions. This suggests that the dorsal ACG area is optimally located to mediate between

sensory input and behavioral responses, particularly information coming from different hemispheres: The right SPL, already known to operate in visual attention, and the left STG/MTG likely feed relevant information upstream. These observations further highlight the often overlooked role of the ACG in allocation of cross-modal attention when cognitive demands are high.

In sum, our results delineate the neural network that interacts with traditional speech processing regions under high-cognitive load conditions. This network includes control mechanisms that lie outside the typical ventral and dorsal language streams, suggesting flexible interactions between sensory processes and top-down attention modulations. The increase in processing demands led to poorer performance on the speech task and decreased activity in speech processing brain regions, while recruitment of medial frontal structures became stronger and varied as a function of individual differences. This suggests a critical role for medial frontal regions in the control of cross-modal divided attention.

Acknowledgements

This study was supported in part by an ESRC grant to S. L. Mattys (RES-062-23-2746). We thank Tatjana Zimasa, Francesca Mandino, and Kris Farrant for help with testing and data analysis.

References

Baker E, Blumstein SE, and Goodglass H (1981). Interaction between phonological and semantic factors in auditory comprehension. *Neuropsychologia*, *19*, 1-15.

Barton M, Marecek R, Rektor I, Filip P, Janousova E, and Mikl M. (2015). Sensitivity of PPI analysis to differences in noise reduction strategies. *Journal of Neuroscience Methods*, *253*, 218-232.

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. arXiv preprint arXiv:1506.04967.

Beckmann C, Jenkinson M, and Smith SM. (2003). General multi-level linear modelling for group analysis in FMRI. *Neuroimage*, *20*, 1052-1063.

Beckmann M, Johansen-Berg H, and Rushworth MF. (2009). Connectivity-based parcellation of human cingulate cortex and its relation to functional specialization. *Journal of Neuroscience*, *29*, 1175-1190.

Botvinick MM. (2007). Conflict monitoring and decision making: Reconciling two perspectives on anterior cingulate function. *Cognitive, Affective & Behavioral Neuroscience*, *7*, 356-366.

Botvinick MM, Braver TS, Barch DM, Carter CS, and Cohen JD. (2001). Conflict Monitoring and Cognitive Control. *Psychological Review, 108,* 624-652.

Botvinick MM, Cohen JD, and Carter CS. (2004). Conflict monitoring and anterior cingulate cortex: an update. *Trends in Cognitive Sciences*, *8*, 539-546.

Brungart D, Iyer N, Thompson ER, and Simpson BD. (2013). Interactions between listening effort and masker type on the energetic and informational masking of speech stimuli. *Proceedings of Meetings on Acoustics*, *19 (1)*, 060146.

Catani M, Howard RJ, Pajevic S, and Jones DK. (2002). Virtual in vivo interactive dissection of white matter fasciculi in the human brain. *Neuroimage*, *17*, 77-94.

Davis MH, and Johnsrude IS. (2003). Hierarchical processing in spoken language comprehension. *Journal of Neuroscience*, *23*, 3423-3431.

Downar J, Crawley AP, Mikulis DJ, and Davis KD. (2001). The effect of task relevance on the cortical response to changes in visual and auditory stimuli: An event-related fMRI study. *NeuroImage*, *14*, 1256-1267.

Ebitz RB, and Hayden BY. (2016). Dorsal anterior cingulate: A Rorschach test for cognitive neuroscience. *Nature Neuroscience*, *19*, 1278-1279.

Eckert, MA, Menon, V, Walczak A, Ahlstrom J, Denslow S, Horwitz A, & Dubno JR. (2009). At the heart of the ventral attention system: the right anterior insula. *Human Brain Mapping*, *30*, 2530-2541.

Erb J, & Obleser J (2013). Upregulation of cognitive control networks in older adults' speech comprehension. *Frontiers in Systems Neuroscience*, *7,* 116.

Francis AL. (2010). Improved segregation of simultaneous talkers differentially affects perceptual and cognitive capacity demands for recognizing speech in competing speech. *Attention, Perception, & Psychophysics*, *72*, 501-516.

Friederici, A. (2011). The brain basis of language processing: From structure to function. *Physiological Review, 91*, 1357-1392.

Friston KJ, Buechel C, Fink GR, Morris J, Rolls E, and Dolan RJ. (1997). Psychophysiological and modulatory interactions in neuroimaging. *Neuroimage*, *6*, 218-229.

Friston KJ, Josephs O, Rees G, and Turner R. (1998). Nonlinear event-related responses in fMRI.

Magnetic resoncance in medicine, *39*, 41-52.

Gitelman DR, Penny WD, Ashburner J, and Friston KJ. (2003). Modeling regional and psychophysiologic interactions in fMRI: The importance of hemodynamic deconvolution. *NeuroImage*, *19*, 200-207.

Grill-Spector K. (2003). The neural basis of object perception. *Current Opinion in Neurobiology*, *13*, 159-166.

Grill-Spector K, Kourtzi Z, and Kanwisher N. (2001). The lateral occipital complex and its role in object recognition. *Vision Research*, *41*, 1409-1422.

Herath P, Klingberg T, Young J, Amunts K, and Roland P. (2001). Neural correlates of dual task interference can be dissociated from those of divided attention: An fMRI study. *Cerebral cortex*, *11*, 796-805.

Hickok G, and Poeppel D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, *8*, 393-402.

Hopfinger JB, Buonocore MH, and Mangun GR. (2000). The Neural Mechanisms of TopDown Attentional Control, *3*. doi:10.1038/72999

Hutchison ER, Blumstein SE, and Myers EB (2008). An event-related fMRI investigation of voice-onset time discrimination. *Neuroimage*, *40*, 342-352.

Hymers M, Prendergast G, Liu C, Schulze A, Young ML, Wastling SJ, … Millman RE. (2015). Neural mechanisms underlying song and speech perception can be differentiated using an illusory percept. *NeuroImage*, *108*, 225-233.

Iidaka T, Anderson ND, Kapur S, Cabeza R, and Craik FI. (2000). The effect of divided attention on encoding and retrieval in episodic memory revealed by positron emission tomography. *Journal of Cognitive Neuroscience*, *12*, 267-280.

Jenkinson M, Bannister. P., Brady M, and Smith S. (2002). Improved optimisation for the robust and accurate linear registration and motion correction of brain images. *Neuroimage*, *17*, 825-841.

Joanisse MF, and Gati JS. (2003). Overlapping neural regions for processing rapid temporal cues in speech and nonspeech signals. *NeuroImage*, *19*, 64-79.

Johnson JA, Strafella AP, and Zatorre RJ. (2007). The role of the dorsolateral prefrontal cortex in bimodal divided attention: Two transcranial magnetic stimulation studies. *Journal of Cognitive Neuroscience*, *19*, 907-920.

Johnson JA, and Zatorre RJ. (2005). Attention to simultaneous unrelated auditory and visual events: Behavioral and neural correlates. *Cerebral Cortex*, *15*, 1609-1620.

Johnson JA, and Zatorre RJ. (2006). Neural substrates for dividing and focusing attention between simultaneous auditory and visual events. *NeuroImage*, *31*, 1673-1681.

Loose R, Kaufmann C, Auer DP, and Lange KW. (2003). Human prefrontal and sensory cortical activity during divided attention tasks. *Human Brain Mapping*, *18*, 249-259.

Macdonald JSP, and Lavie N. (2011). Visual perceptual load induces inattentional deafness. *Attention, Perception, & Psychophysics*, *73*, 1780-1789.

MacDonald AW, Cohen JD, Stenger VA, and Carter CS. (2000). Dissociating the role of the dorsolateral prefrontal and anterior cingulate cortex in cognitive control. *Science*, *288*, 1835-1838.

Machielsen WCM, Rombouts S a RB, Barkhof F, Scheltens P, and Witter MP. (2000). fMRI of visual encoding: Reproducibility of activation. *Human Brain Mapping*, *9*, 156-164.

Mattys SL, Barden K, and Samuel AG. (2014). Extrinsic cognitive load impairs low-level speech

perception. *Psychonomic Bulletin & Review*, *21*, 748-754.

Mattys SL, and Palmer SD. (2015). Divided attention disrupts perceptual encoding during speech recognition. *The Journal of the Acoustical Society of America*, *137*, 1464-1472.

Mattys SL, and Wiget L. (2011). Effects of cognitive load on speech recognition. *Journal of Memory and Language*, *65*, 145-160.

Miller, E. K. & Cohen, J. D. (2001) An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, *24*, 167-202.

Molenberghs P, Mesulam MM, Peeters R, and Vandenberghe RRC. (2007). Remapping attentional priorities: Differential contribution of superior parietal lobule and intraparietal sulcus. *Cerebral Cortex*, *17*, 2703-2712.

Molloy K, Griffiths TD, Chait M, and Lavie N. (2015). Behavioral/Cognitive Inattentional Deafness: Visual Load Leads to Time-Specific Suppression of Auditory Evoked Responses. *Journal of Neuroscience*, *35*, 16046-16054.

Mueller K, Mildner T, Fritz T, Lepsien J, Schwarzbauer C, Schroeter ML, and Möller HE. (2011). Investigating brain response to music: A comparison of different fMRI acquisition schemes. *NeuroImage*, *54*, 337-343.

Myers EB, Blumstein SE, Walsh E, and Eliassen J. (2009). Inferior frontal regions underlie the perception of phonetic category invariance. *Psychological Science*, *20*, 895-903.

Navon D (1984). Resources - A theoretical soup stone? *Psychological Review*, *91*, 216-234.

Navon D, and Gopher D (1979). On the economy of the human processing system. *Psychological Review*, *86*, 214-255.

O'Reilly JX, Schüffelgen U, Cuell SF, Behrens TEJ, Mars RB, and Rushworth MFS. (2013).

Dissociable effects of surprise and model update in parietal and anterior cingulate cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *110*, E3660-E3669.

Obleser J, Wise RJS, Dresner MA, and Scott SK. (2007). Functional integration across brain regions improves speech perception under adverse listening conditions. *Journal of Neuroscience*, *27*, 2283-2289.

Peelle J. (2014). Methodological challenges and solutions in auditory functional magnetic resonance imaging. *Frontiers in Neuroscience*, *8*, 253.

Picard N, and Strick PL. (1996). Motor areas of the median wall: a review of their location and functional activation. *Cerebral Cortex*, *6*, 342-353.

Price CJ. (2012). A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *NeuroImage*, *62*, 816-847.

Raveh D, and Lavie N. (2015). Load-induced inattentional deafness. *Attention, Perception, & Psychophysics*, *77*, 483-492.

Rinne T, Balk MH, Koistinen S, Autti T, Alho K, and Sams M. (2008). Auditory selective attention modulates activation of human inferior colliculus. Jo*urnal of Neurophysiology*, *100*, 3323-3327.

Rushworth MFS, and Behrens TEJ. (2008). Choice, uncertainty and value in prefrontal and cingulate cortex. *Nature Neuroscience*, *11*, 389-97.

Rypma B, and D'Esposito M. (1999). The roles of prefrontal brain regions in components of working memory: effects of memory load and individual differences. *Proceedings of the National Academy of Sciences of the United States of America*, *96*, 6558-6563.

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

Salo E, Rinne T, Salonen O, and Alho K. (2015). Brain activations during bimodal dual tasks depend on the nature and combination of component tasks. *Frontiers in Human Neuroscience*, *9*, 102.

Schwarzbauer C, Davis MH, Rodd JM, and Johnsrude I. (2006). Interleaved silent steady state (ISSS) imaging: A new sparse imaging method applied to auditory fMRI. *NeuroImage*, *29*, 774-782.

Scott SK, Blank CC, Rosen S, and Wise RJS. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain*, *123*, 2400-2406.

Shenhav A, Botvinick MM, and Cohen J. (2013). The expected value of control: An integrative theory of anterior cingulate cortex function. *Neuron*, *79*, 217-240.

Shomstein S, and Yantis S. (2004). Control of attention shifts between vision and audition in human cortex. *Journal of Neuroscience*, *24*, 10702-10706.

Smith S. (2002). Fast robust automated brain extraction. *Human Brain Mapping*, *17*, 143-155.

Sörqvist P, Stenfelt S, and Rönnberg J. (2012). Working Memory Capacity and Visual-Verbal Cognitive Load Modulate Auditory-Sensory Gating in the Brainstem: Toward a Unified View of Attention. *Journal of Cognitive Neuroscience*, *24*, 2147-2154.

Vaden KI, Teubner-Rhodes S, Ahlstrom JB, Dubno JR, & Eckert MA (2017). Cingulo-opercular activity affects incidental memory encoding for speech in noise. *NeuroImage, 157*, 381-387.

Vaidya CJ, Zhao M, Desmond JE, and Gabrieli JDE. (2002). Evidence for cortical encoding specificity in episodic memory: Memory-induced re-activation of picture processing areas.

Neuropsychologia, *40*, 2136-2143.

Wei P, Szameitat AJ, Müller HJ, Schubert T, and Zhou X. (2013). The neural correlates of perceptual load induced attentional selection: An fMRI study. *Neuroscience*, *250*, 372-380.

Wojciulik E, and Kanwisher N. (1999). The generality of parietal involvement in visual attention. *Neuron*, *23*, 747-764.

Woolrich MW, Behrens TEJ, Beckmann CF, Jenkinson M, and Smith SM. (2004). Multi-level linear modelling for FMRI group analysis using Bayesian inference. *Neuroimage*, *21*, 1732-1747.

Worsley K. (2001). Testing for signals with unknown location and scale in a $\chi 2$ random field, with an application to fMRI. *Advances in Applied Probability*, *33*, 773-793.

Table 1. Peak voxels of significant activation clusters in the whole brain and ROI analyses.

			MNI coordinates			
Contrast	Atlas labels	*Z* value	X	y	z	voxels
High load > low load	R LOC	4.69	40	-86	-6	2411
	L OP/LOC	5.15	-38	-90	6	1737
	R PaCG	3.81	2	16	44	84
	(R SPL)	3.75	28	-52	48	60
	(L ACG)	3.37	-8	28	26	97
Low load > high load	L LG	4.16	-8	-78	-2	201
	L pSTG/pMTG	3.92	-60	-38	4	155

Notes. LOC: Lateral Occipital Cortex; SPL: Superior Parietal Lobule; OP: Occipital Pole; ACG: Anterior Cingulate Gyrus; PaCG: Paracingulate Gyrus; LG: Lingual Gyrus; pSTG: posterior Superior Temporal Gyrus; pMTG: posterior Middle Temporal Gyrus; L: left; R: right. Whole brain and ROI analyses were thresholded at $Z = 3.1$, $p = .05$ (cluster corrected). In parentheses are additional clusters found in a whole brain analysis at a more lenient threshold ($Z = 2.3$, $p = .05$).

Table 2. Peak voxels of the significant clusters in the PPI analysis.

		MNI coordinates		
Atlas labels	Z value	x	Y	z
R ACG	3.37	14	38	16
R PaCG	3.42	16	50	16
R ACG	3.27	4	34	22
L PaCG	2.94	-16	52	8
L PaCG	2.73	-12	42	16
L ACG	2.37	-8	32	20
L ACG	2.92	0	34	22

Note. The seed region of the PPI analyses was a 6mm sphere centred at coordinates x = -60, y = -38, z = 4 in L pSTG/pMTG. See Table 1 for labels.

Figure Captions

Figure 1: Schematic representation of the divided-attention tasks and acquisition protocol for a single trial. Each trial started with a rest period of 6 s during which a fixation cross was displayed (not shown). The visual and auditory stimuli were then presented simultaneously. In the low-load condition, the four consecutive visual images were the same whereas, in the high-load condition, they were different. Data acquisition started at the end of stimulus presentation. Data were acquired in four volumes of 2 s each (8 s total). A picture of headphones appeared after two volumes, which prompted participants to decide if the two syllables were the same or different (auditory task). At the end of the third volume, a visual image appeared and participants had to decide whether the image had been presented earlier in the trial (visual task).

Figure 2: Behavioral performance in the phoneme discrimination task (left panel) and in the visual task (right panel) as a function of Load (low load vs. high load) and Auditory Pair (same vs. different).

Figure 3: BOLD contrasts (group mean effect) during the dual task. High-load > low-load contrasts are shown in red. Low-load > high-load contrasts are shown in blue. The top panel shows the results of the ROI analysis within the left frontal lobe (cluster corrected, $Z = 3.1$, $p = .05$). The top bar graph shows the average percentage signal change across the significant voxels of the top panel. The bottom panel shows activity in the lingual gyri and the posterior middle and superior temporal gyri (in blue) (whole brain cluster corrected, $Z = 3.1$, $p = .05$). The bottom bar graph shows the average percentage signal change across the significant voxels of the posterior temporal cluster.

Figure 4: Scatterplots of the average percent signal change for the contrast high-load > low-load in the paracingulate and anterior cingulate cluster with the average percentage signal change in the

pSTG/pMTG cluster (left panel) and the average behavioral performance in the visual task (right panel).

Figure 5: Brain clusters showing psycho-physiological interaction with the seed region in pSTG/pMTG. Cluster thresholding was performed using thresholds of $Z > 2.0$, $p < .05$ and are displayed for voxels with Z values > 2.3.

Figure 1

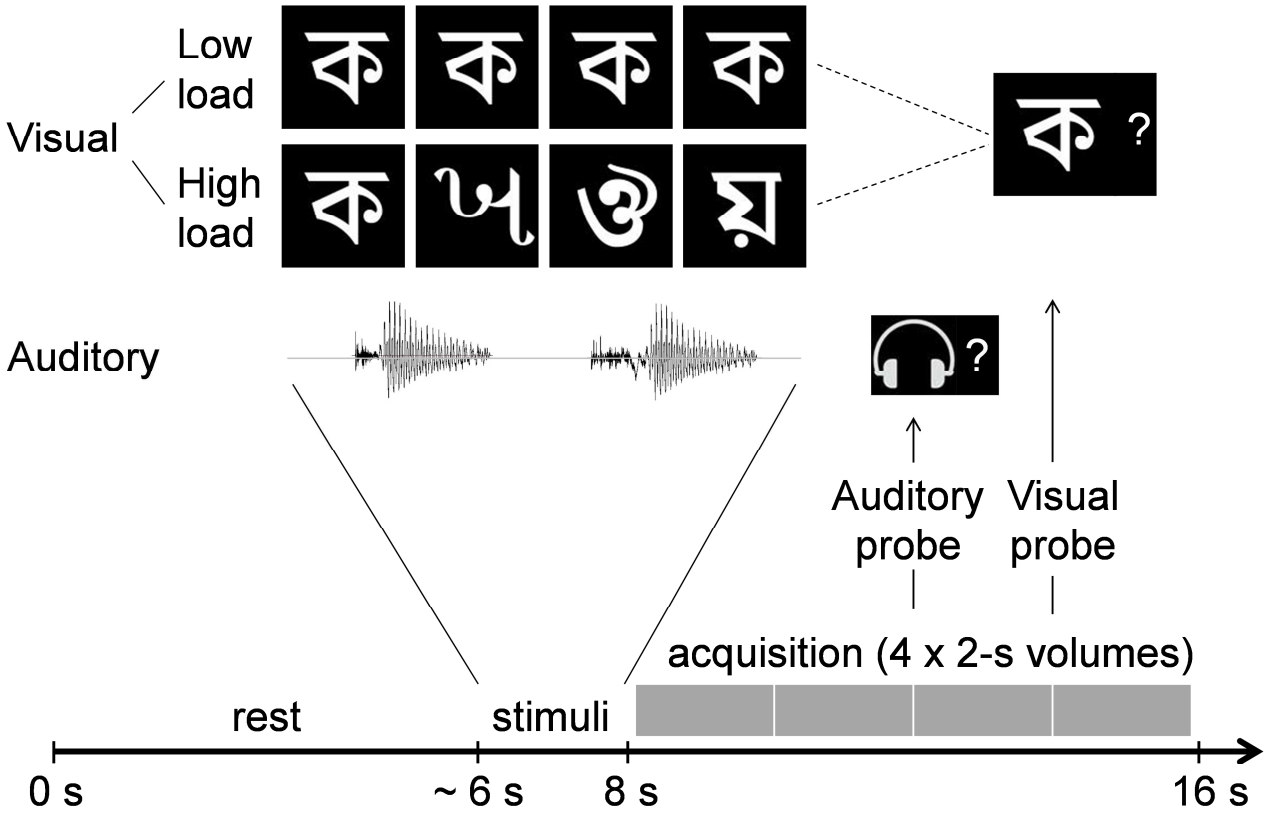


Figure 2

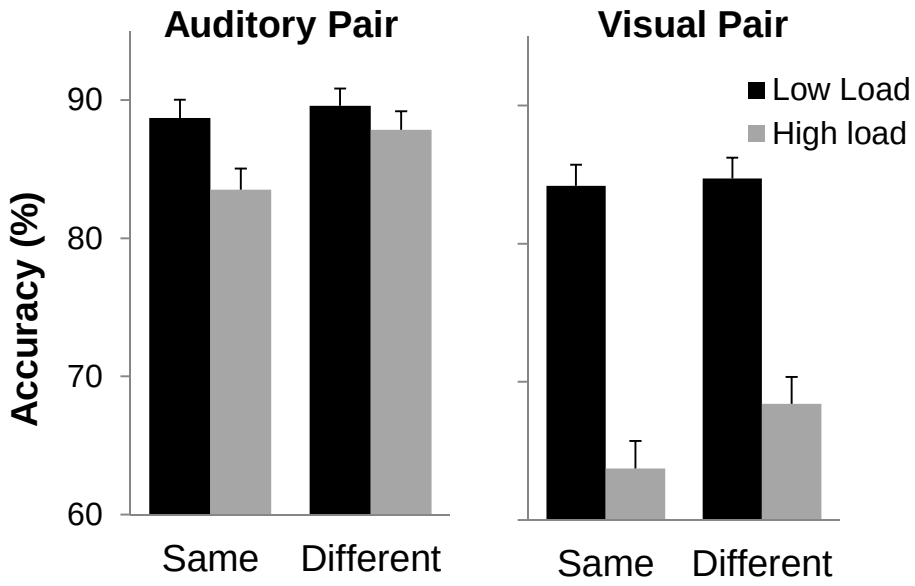


Figure 3

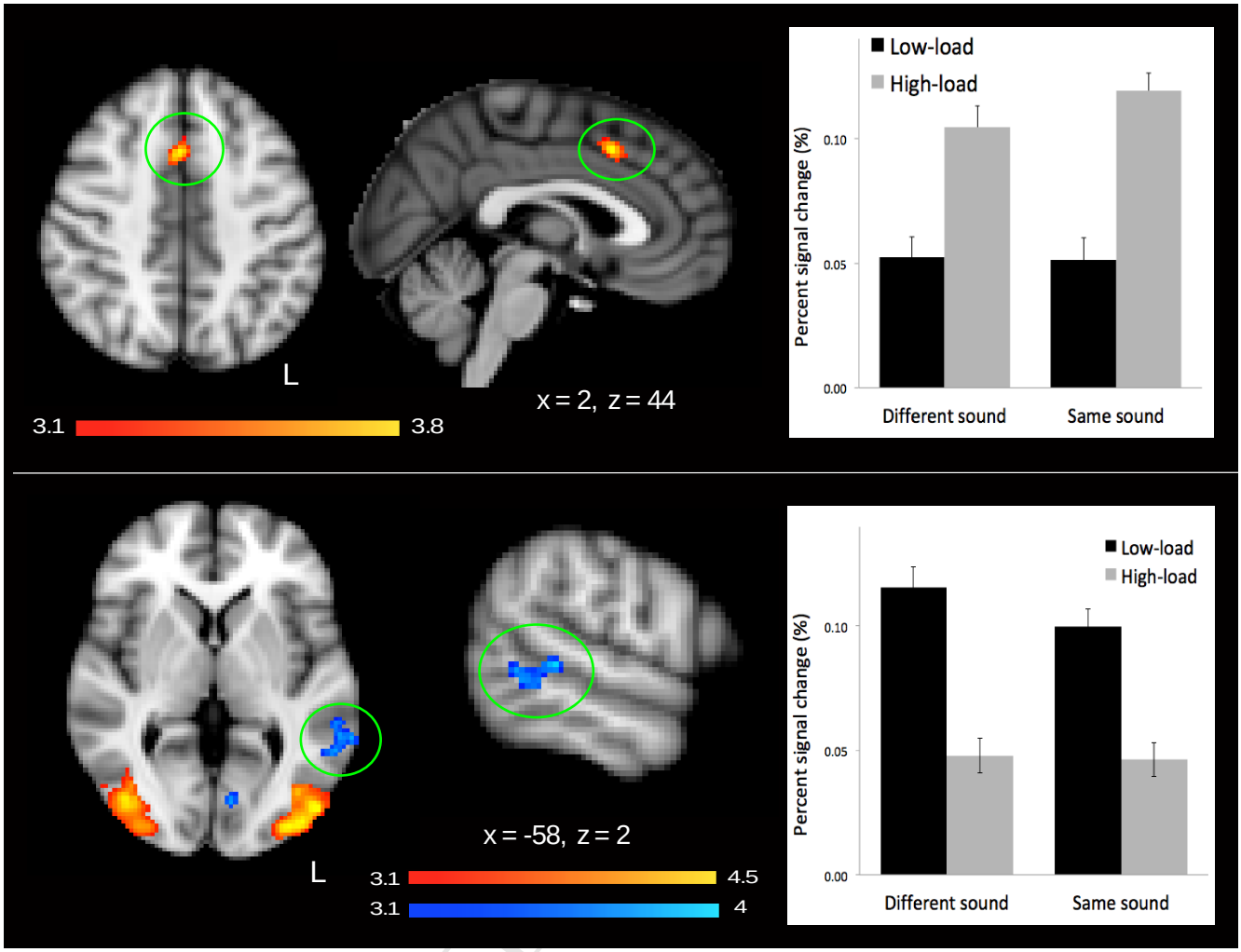


Figure 4

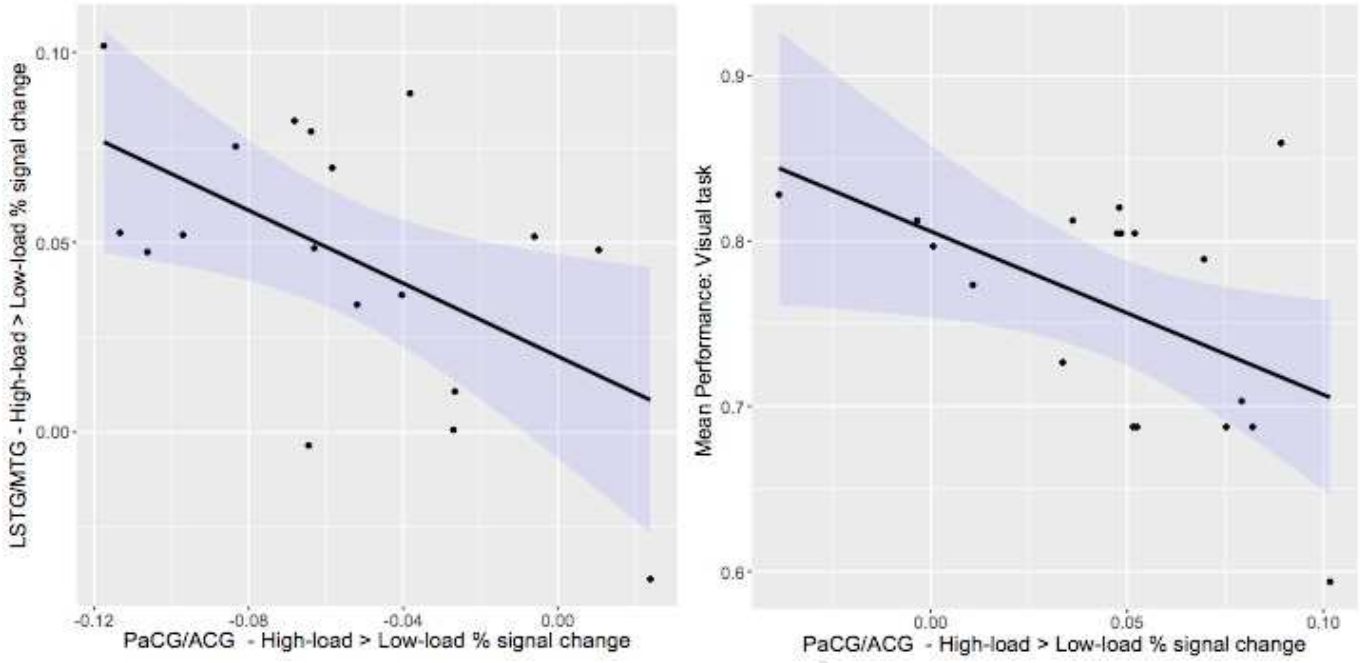


Figure 5