8-1-2012

Behavioral and fMRI evidence that cognitive ability modulates the effect of semantic context on speech intelligibility

Adriana A Zekveld

Mary Rudner

Ingrid Johnsrude

Dirk J Heslenfeld

Jerker Rönnberg

Behavioural and fMRI evidence that cognitive ability modulates the effect of semantic context on speech intelligibility.

Adriana A. Zekveld[a,b,c], Mary Rudner[a,b], Ingrid S. Johnsrude[a,b,d], Dirk J. Heslenfeld[e], Jerker Rönnberg[a,b]

[a]Department of Behavioural Sciences and Learning, Linköping University, SE-581 83 Linköping, Sweden

[b]Linnaeus Centre HEAD, The Swedish Institute for Disability Research, Linköping and Örebro Universities, SE-581 83 Linköping, Sweden

[c]Audiology / ENT & EMGO+ Institute for Health and Care Research, VU University medical center, De Boelelaan 1118, 1081 HZ, Amsterdam, The Netherlands

[d]Department of Psychology, Queen's University, Kingston, Humphrey Hall, 62 Arch Street, Ontario, Canada, K7L 3N6

[e]Department of Psychology, VU University, Van der Boechorststraat 1, 1081 BT Amsterdam, The Netherlands

Email addresses of the authors:

adriana.zekveld@liu.se; mary.rudner@liu.se; ingrid.johnsrude@queensu.ca; dirk@psy.vu.nl; jerker.ronnberg@liu.se

Correspondence concerning this article should be addressed to Adriana A. Zekveld

Address for correspondence:

ENT /Audiology and the EMGO+ Institute for Health and Care Research,

VU University medical center, P.O. Box 7057,

1007 MB Amsterdam, The Netherlands.

Telephone: +3120 444 0952, Fax: +3120 444 2033,

Email: adriana.zekveld@liu.se

Abstract

Text cues facilitate sentence perception when they are semantically related to the speech (Zekveld et al., 2011a). In this study, semantically related and unrelated cues preceding sentences evoked more activation in middle temporal gyrus (MTG) and inferior frontal gyrus (IFG) than non-word cues, regardless of acoustic quality (speech in noise or speech in quiet). Larger verbal working memory (WM) capacity (reading span) was associated with greater intelligibility benefit obtained from related cues, with less speech-related activation in the left superior temporal gyrus and left anterior IFG, and with more activation in right medial frontal cortex for related versus unrelated cues. Better ability to comprehend masked text was associated with greater ability to disregard unrelated cues, and with more activation in left angular gyrus (AG). We conclude that individual differences in cognitive abilities are related to activation in a speech-sensitive network including left MTG, IFG and AG during cued speech perception.

Keywords: fMRI, spoken sentence perception, text cues, working memory

1. Introduction

Speech understanding is a complex process depending on both stimulus-driven and knowledge-driven processes. Audibility influences the relative importance of these processes (c.f., Plomp, 2002; Rönnberg, Rudner, Lunner, & Zekveld, 2010). When hearing is challenged by interfering noise and/or hearing impairment, contextual information, prior knowledge and experience play a greater role in driving perception (Rönnberg, 2003; Rönnberg, Rudner, Foo, & Lunner, 2008; Wingfield & Tun, 2007; Norris & McQueen, 2008; Miller, Heise, & Lichten, 1951; Miller & Isard, 1963, Tesink et al., 2009; Nakano, Saron, & Swaab, 2010; Hagoort & van Berkum, 2007; Drager & Reichle, 2001).

Previous research (Rönnberg et al., 2008; Akeroyd, 2008; Arlinger, Lunner, Lyxell, & Pichora-Fuller, 2009; Rudner, Rönnberg, & Lunner, 2011, Engle, 2002) indicates that speech perception, especially in challenging listening conditions, is associated with working memory (WM) capacity as measured with the reading span test (RSpan, Daneman & Carpenter, 1980). It also relates to the ability to read partly masked everyday sentences as assessed by the text reception threshold test (TRT; Zekveld, George, Kramer, Goverts & Houtgast, 2007).

The relation between WM and speech perception likely reflects processes including the integration of incoming information with the information stored in memory, and the evaluation of hypotheses based on the results of earlier processes against incoming auditory information (Kane & Engle, 2000). According to the Ease of Language Understanding (ELU) framework (Rönnberg, 2003), such processes are increasingly recruited when stimuli are degraded such that the stimulus-driven linguistic input cannot be automatically and implicitly matched to stored phonological representations in long-term memory. Such failures to match, or 'mismatches' (see Rönnberg et al., 2008), require explicit, knowledge-driven processing in order to be resolved. In contrast, clearly audible stimulus-driven linguistic input may not

involve WM as much, since it is more automatically and implicitly bound to stored representations.

In the Text Reception Threshold test (Zekveld et al., 2007), subjects are asked to read sentences that are presented as text partly masked by a bar pattern mask. This test was developed to assess cognitive abilities that may also be relevant for speech perception in noise. Because stimuli are presented as text, only non-auditory abilities relevant for understanding partly masked linguistic information are addressed, so that performance on the test specifically reflects functions that are independent of auditory processes. These functions likely include the ability to form hypotheses about the meaning of a sentence based on sentence fragments. Performance on this test correlates with the intelligibility of speech in noise (Kramer et al., 2009; Besser, Zekveld, Kramer, Rönnberg, & Festen, 2012).

In a previous behavioural study (Zekveld et al., 2011a), we assessed the association between WM, TRT, and the ability to capitalise on contextual cues during speech understanding in noise. We presented 3-word text cues immediately prior to spoken sentences that were presented in noise at a range of signal-to-noise ratios (SNRs). Text cues were either *related* or *unrelated* to the meaning of the sentence. *Non-word* cues served as a neutral baseline. We assumed that two aspects of this test would particularly recruit processes dependent on working memory. First, the text cues have to be stored in memory in order to integrate their semantic information with the spoken sentence content presented after disappearance of the cues. Second, subjects were aware of the fact that both related and unrelated cues were presented and this might evoke explicit mental testing of the association between cues and sentence. The test paradigm was further assumed to evoke linguistic processes that are also relevant for performing the TRT test, i.e., inferring the probable speech content from available linguistic information (cues).

The behavioral data obtained by Zekveld et al. (2011a) indicated that related cues improved perception relative to non-word and unrelated cues, with larger benefit obtained at lower SNRs. Furthermore, at low SNRs, unrelated cues slightly impaired perception relative to non-word cues. Intriguingly, higher RSpan and TRT scores were associated with better sentence perception performance after unrelated cues, at least at low SNRs. This suggests that individuals with stronger cognitive abilities were better able to disregard unhelpful cues (c.f. Sörqvist, Ljungberg and Ljung 2010). Recently, Besser et al. (2012) compared the relation between RSpan, TRT and speech perception performance. They showed that TRT performance is less age-dependent than RSpan capacity, and that introducing a WM component in the TRT task did not appreciably change its ability to predict speech perception in noise. Hence, they concluded that RSpan and TRT tap into different cognitive processes relevant for speech perception in noise.

Neuroimaging studies have examined the relation between cognitive abilities (e.g., Just, Carpenter, & Miyake, 2003), speech perception and brain activity (e.g., Mason & Just, 2007; Osaka et al., 2003). These studies indicate that RSpan can explain individual differences in the efficiency of processes (attentional or lexical) relevant for understanding degraded speech.

The present study was designed to obtain more insight into the neural correlates of the interaction between semantic context and speech perception, and to assess its association with two cognitive tasks, both thought to tap relevant aspects of working memory and linguistic abilities. If neural activation related to using context to support speech comprehension is higher in individuals with larger WM capacity, this would suggest that these individuals may recruit additional semantic processing networks whereas the opposite relation, with less activation in high-capacity individuals, would suggest more efficient processing of context and speech (c.f., the 'neural efficiency hypothesis'; Neubauer & Fink, 2009). Regarding the

relation between TRT performance and cognitive processing load, Zekveld, Kramer, and Festen (2011b) showed that individuals with good TRT performance experienced *more* processing load during speech perception in noise, as measured by examination of the pupil response.

In the current study, text cues served as semantically related, unrelated or neutral context for speech presented in quiet or noise. Work to date has shown that left-frontal regions including the IFG are sensitive to sentence context, especially when speech intelligibility is low (Obleser, Wise, Dresner & Kotz, 2007; Obleser & Kotz, 2010), possible due to involvement of this area in semantic and contextual computation (Binder & Desai, 2011).

Consistent with Obleser and Kotz (2010), we expected that sentences preceded by semantically related cues would evoke less activation in left IFG compared to sentences preceded by unrelated and nonword cues. Additionally, we predicted an interaction between the effect of the cues and acoustic quality (speech in noise or speech in quiet), with more extended activation in fronto-temporal regions for related cues compared to unrelated cues (Obleser & Kotz, 2010). We also expected that in the present study RSpan and TRT would be differentially related to neural activation during sentence processing, given Besser and colleagues' (2012) observation that these two measures are associated with different aspects of speech comprehension. Based on the 'neural efficiency hypothesis' (Neubauer & Fink, 2009) and the predictions of the ELU framework, we expected that larger WM capacity will be associated with less activation in left IFG areas. Based on Zekveld et al. (2011b), we predicted greater activity in left IFG regions, reflecting greater processing load, in listeners with better TRT performance.

2. Methods

2.1 Participants

Eighteen normal-hearing young adults (9 women and 9 men) participated. None of them had taken part in previous studies using a similar test paradigm. Ages ranged from 20 to 29 yrs, with a mean age of 23.6 yrs ($SD = 2.6$ yrs). They were recruited from among students and employees of VU University and VU University Medical Centre, and via advertisements. All participants were native Dutch speakers who reported having no hearing problems. They all had normal or corrected to normal vision (correction between -6 to +4 diopters when wearing glasses and able to read text with font size 7 or lower as screened with a near vision test equivalent to a visual acuity Snellen chart; Bailey & Lovie, 1980). All were right-handed on the Dutch "Classificatie van links- en rechtshandige proefpersonen" ("Classification of left- and right-handed subjects") (Van Strien, 1992). The exclusion criteria were the following: having pure-tone hearing threshold(s) exceeding 20 dB HL at the octave frequencies between 500-8000 Hz, dyslexia or other reading problems, claustrophobia, epilepsy, having a history of a neurological or psychiatric disease, having metal in the body that would preclude safety of an MRI scan, and being pregnant or on medication. All participants provided written informed consent in accordance with the Ethics Committee of the VU University Medical Center.

2.2 Procedure

Participants were tested in two sessions that took place on different days in all but two cases; these two participants were tested in both sessions on the same day and had several hours to rest and recover in between the sessions. On average, session 2 was 13 days after session 1. The first session was behavioural and included tests to work out an appropriate SNR for the imaging study. The second session comprised the fMRI study. We will first summarize the procedures, and then describe them in detail.

2.2.1 Summary of procedures

Session 1 took place in a sound-attenuating room. We obtained a pure-tone audiogram and checked the participants' near vision. Then participants filled in the handedness questionnaire. We additionally measured uncued Speech Reception Threshold (SRT; Plomp & Mimpen, 1979; see below). After a practice SRT test, we determined the speech-to-noise ratio (SNR) resulting in the entirely correct perception of 29% of the sentences. Results of our previous study (Zekveld et al., 2011a) showed that the influence of semantic context depended on the acoustic quality of the speech. In the current study, we aimed to control for interindividual differences in baseline intelligibility level (i.e., sentence intelligibility without context) by individually adapting the SNR at which the stimuli were presented. In Zekveld et al. (2011a) we observed that at 29% baseline intelligibility, prior semantic context has the largest effect on sentence intelligibility. Therefore, this individually estimated threshold was used as the SNR in Session 2 (see below). We also determined the Text Reception Threshold (TRT; Zekveld et al, 2007; see below). Subjects performed one practice and three regular TRT tests, and we used the TRT averaged over the three tests in the analysis. Finally, the Reading Span test (Daneman & Carpenter; 1980; Besser et al., 2012; see below) was performed to determine working memory capacity. The duration of Session 1 was 1 hour.

In Session 2, the cued speech perception test was performed inside the scanner. The duration of Session 2 was 1.5 hours.

2.2.2 Detailed procedures

2.2.2.1 Speech Reception Threshold test

In each SRT test (for details, see Plomp & Mimpen, 1979), 26 Dutch sentences were presented in background noise. Sentences were selected from the set of 507 sentences as developed by Versfeld, Daalder, Festen and Houtgast (2000) and recorded at 44100 Hz / 16

bits by a female speaker. An example sentence (translated into English) is "the shop is within walking distance" (see Versfeld et al., 2000, for more details). Stationary noise with a long-term average spectrum of the speech served as masker. For each sentence, the presentation of the noise started 500 msec before speech onset and noise-offset was 500 msec after speech offset. Participants were asked to report each sentence aloud. The trial was correct if every word in the sentence was reproduced completely correctly (sentence scoring). The SNR was varied by adapting both the speech and the noise levels while keeping the overall intensity fixed at 70 dB SPL. A 1-up-2-down staircase adaptive procedure with a step-size of 2 dB SNR was applied, estimating the SNR required for perceiving 29% of the presented sentences entirely correctly (Levitt, 1971). The SRT was the mean SNR of sentences 5 to 27; the 27th sentence was not actually presented, but the SNR at which this sentence would have been presented can be calculated from the response to the 26th sentence. Lower SRTs indicate better performance. Sentences did not overlap with those presented in the other tests.

2.2.2.2 Text Reception Threshold (TRT) test

The TRT test is a text-based test conceptually equivalent to the SRT test: it measures the ability to perceive degraded verbal information. In the TRT test, partly masked printed sentences (Versfeld et al., 2000) were presented (Zekveld et al., 2007). The sentences were different to those presented in the other tests. The text was masked with a bar pattern (see Zekveld et al., 2007). The field background color was white, text color was red, and the color of the mask was black. At the start of each trial, the mask became visible and the text appeared 'behind' it in a word-by-word fashion. Display-onset of each word in the sentence was equal to the timing of the start of the utterance of each word in the corresponding audio file (Versfeld et al., 2000). The average duration of the utterance of the words in these audio files was 281 msec, ranging from 44 to 854 msec. All words remained on the screen for 3500

ms after the sentence was complete. Subjects were asked to read the sentences aloud. The experimenter scored whether the sentences were reproduced entirely correctly. A 1-up-1-down adaptive procedure with a step-size of 6% was applied, targeting the percentage of unmasked text required to read 50% of the sentences entirely correctly. The TRT was the mean percentage of unmasked text of sentence 5 to 27, lower TRTs indicate better performance.

2.2.2.3 Reading Span test

In the Reading Span Test (Daneman & Carpenter, 1980), 5-word sentences are presented visually. The Dutch sentence materials were developed (see Besser et al., 2012) to be equivalent to the Swedish version described by Andersson, Lyxell, Rönnberg, and Spens (2001; see also Rönnberg, Arlinger, Lyxell, & Kinnefors, 1989) in turn based on an English version (Baddeley, Nimmo-Smith & Brereton, 1985). Our previous study (Zekveld et al., 2011) showed that this test is associated with the benefit obtained from contextual cues during speech perception. Half of the sentences are nonsense (e.g., the table sings a song); the other half are coherent (e.g., the friend told a story). First, three sets of three sentences are presented, followed by three sets of four sentences, three sets of five sentences, and three sets of six sentences. Immediately after each sentence, participants verbally indicate whether the sentence makes sense or not. After each set of sentences, participants are prompted to recall, in serial order, either the first or the last word of each sentence. The experimenter records the total number of words correctly recalled.

2.3 Imaging Session: Cued Speech Perception

2.3.1 Stimuli and Conditions

To create the auditory stimuli, we employed a two-factor within-subjects factorial design, crossing cue type (related cues, unrelated cues, or non-word cues) with noise level (speech in quiet or speech in noise). Additionally, three types of baseline stimuli were presented: noise alone preceded by either nonword or unrelated cues, and silence preceded by nonword cues. An overview of the conditions is presented in Table 1.

We used 243 sentences with durations between 1.32 s and 2.70 s (mean =1.84 s) selected from the same set of (but not overlapping with) sentences used in the SRT test (Versfeld et al., 2000). Sentences were grouped into nine lists of 27 items; for each subject, six of these nine sentence lists were assigned to different stimulus conditions (in a counterbalanced fashion) as follows:

Table 1. Overview of conditions, and an example showing the allocation of sentences and cues to conditions.

Auditory stimuli	Cue type		
	Non-word	Unrelated	Related
SIQ	A_{nw}	C_g	E_e
SIN	B_{nw}	D_h	F_f
Noise alone	$Noise_{nw}$	$Noise_i$	
Silence	$Silence_{nw}$		

Nine lists (A – I) of 27 sentences each were created. The subscripts refer to lists of cues; for sentence-list 'A', the corresponding (i.e., related) cue-list is referred to as '$_a$'; nw = non-word cues. To each individual subject, only 6 sentence lists and 5 cue lists were presented. The allocation of sentence and cue lists to conditions was balanced across subjects. SIQ = Speech in Quiet, SIN = Speech in Noise.

For each participant, two lists of sentences were presented with semantically related cues. Related cues were words that were semantically related to the auditory sentence, and were generated by participants in a previous study (Zekveld et al, 2011a). In that study, 27 participants were presented with 378 text sentences (Versfeld et al., 2000), one at a time, and were asked to generate three words that they thought meaningfully summarized the topic of each sentence (Mäntylä & Nilsson, 1988). Words that occurred in the sentence, parts of compound words occurring in the sentence, or morphological variants of words in the sentence (e.g., the plural form of a noun, or the same verb in another tense) were not valid. For example, for the sentence "It smells nice in the bakery", cue words could be "bread-oven-cakes". We selected the 243 sentences that yielded the most consistently reported cues, determined by counting the number of times participants generated the same individual words for a particular sentence. As a measure of consistency, we calculated the sum of the frequencies for the three most frequently generated words for each sentence. The newly created "most-frequent" three-word cues were new combinations of those 3 individual words that were most often generated for each sentence. We did not expect that creating new combinations of cues would influence the effect of the cues on speech intelligibility. The cues consisted of nouns (59%), verbs (15%), adverbs (14%) and adjectives (12%). For further details see Zekveld et al. (2011a).

Two lists of sentences were presented with semantically unrelated cues. Unrelated cues were related cues for two of the sentence-lists in the stimulus set that were not presented to that particular subject. Unrelated cues did not correspond to any of the three related cues for that sentence, or to any of the words in the sentence itself.

Two lists of sentences were presented with nonword cues. Nonword cues were made by combining three nonwords, created by shuffling the letters of a large set of Dutch words that were randomly selected from the CELEX database (Baayen, Piepenbrock, & Gullikers,

1995). We only used those nonwords that were not pronounceable, and did not contain Dutch words of more than two characters. The distribution of nonword lengths was similar to the distribution of word lengths for the word cues.

For each participant, one of each of the two lists of sentences presented with related, unrelated and nonword cues respectively was presented in quiet (Speech in quiet; SIQ) and the other was presented in stationary noise with a frequency spectrum equal to the long-term average of the speech (Speech in noise; SIN). On SIN trials, the applied SNR was the individually estimated 29% SRT threshold. As in the SRT test, noise onset was 500 ms before sentence onset and noise offset was 500 ms after the end of the sentence.

The remaining three conditions consisted of baseline stimuli: 1) stationary noise with nonword cues; 2) stationary noise with unrelated cues (i.e., cues corresponding to the third list of sentences not presented to that particular subject), and 3) silence with nonword cues. The duration of the noise-alone stimuli and the duration of the interval between the cues and the acquisition of the functional images in the silent condition were chosen such that the mean and variance of the distribution was comparable to that of the speech-in-noise stimuli.

Across participants, each list of 27 sentences was presented in each condition an equal number of times (3 times). Also, across participants, the three lists of unrelated cues were presented an equal number of times (3 times) in each of the three "unrelated" conditions. The presentation order of stimuli from the 9 conditions was pseudo-randomized and interleaved such that each condition was followed an equal number of times (3 times) by each of the 9 conditions. In total, we presented 27 trials in each condition. All stimuli were presented at 70 dB SPL.

Participants were asked to repeat each sentence and were encouraged to make their best guess for sentences that they were not able to understand entirely. They were told that the cues might or might not help them to perceive the sentence. They were instructed to say "Ik

kon deze zin niet verstaan" ("I did not understand the sentence)" when they were not able to perceive any of the words of the speech-in-noise stimuli. For the silent intervals and noise-alone trials, they were instructed to respond "Er was geen zin" ("There was no sentence"). The verbal responses of the participants were captured using the standard intercom system, digitized and stored on a laptop PC located outside the scanner room. Responses were scored offline. We scored and analyzed the number of words reproduced correctly for each sentence.

2.3.2 Imaging Procedure

Brain imaging was performed on a 3 Tesla GE Signa scanner (General Electric Company, Fairfield, CT, USA) equipped with a phased array head coil. We obtained 243 echo-planar imaging (EPI) volumes, consisting of 38 oblique slices each (ascending sampling). The following parameters were used: repetition time (TR) = 16 s, echo time (TE) = 35 ms, volume acquisition time (AT) = 2200 ms, flip angle = 90°, slice thickness = 2.5 mm, slice gap = 0.5 mm, acquisition matrix = 64 x 64 pixels, in-plane resolution = 3.4 mm^2, field of view (FOV) = 220 mm². After the fMRI session, a T1-weighted whole-brain structural scan (MP-RAGE), with a resolution of 1 mm x 0.9 mm x 0.9 mm was acquired.

A PC controlled stimulus presentation. Stimuli were presented using custom-made software based on Delphi 6.0 (Borland Software Corporation, Scotts Valley, CA, USA).

Sounds were presented through a FastTrack M-Audio external sound card (Avid Technology, Inc., Burlington, MA, USA). Stimuli were presented diotically through an MRI-compatible STAX SR-001 MK2 electrostatic headset (STAX Ltd., Saitama, Japan) connected with 25 cm plastic tubes coupled to the ear with foam earplugs to reduce scanner noise. The auditory stimuli were prefiltered in order to correct for the transfer function of the auditory transducer set.

Text stimuli were back-projected onto a screen that was viewed by the participants through an angled mirror positioned on top of the head coil.

The timing of the trials is shown in Figure 1. The sparse sampling paradigm allowed stimulus presentation during the 13.8-sec long quiet periods between scans (Hall et al., 1999). At the beginning of each trial, cues were presented for 2 s on the screen; the auditory stimulus began 50 ms after the cues disappeared from the screen. The sparse sampling procedure prevented the scanner noise and verbal repetition of the sentences from confounding the BOLD response to the following stimulus.
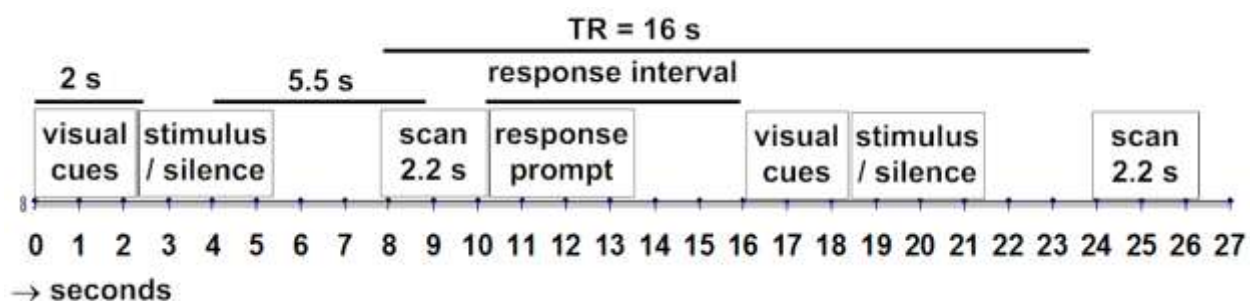

*Figure 1.*Timing of events in the cued speech-perception test.

The onset of the auditory stimuli was jittered relative to the acquisition of the MRI volume (2.2 s) by temporally aligning the midpoint of the stimulus 5.5 s before the midpoint of the image acquisition. After each scan, a response prompt ("herhaal" [repeat]) appeared on the screen for 2 s. Depending on the length of the next auditory stimulus, the response period varied between 6.4 and 7.1 s in order to keep the TR fixed at 16 s. The 243 trials were presented in nine 7.2-minute runs of 27 trials, with three trials from each condition in each run. The scanning session took 65 min in total. Prior to the presentation of the first block of stimuli, participants were given a practice run of 10 trials. At the end of each run, participants were given a short break.

2.4 Preprocessing of the MRI data

Preprocessing and statistical analysis of the MRI data was performed using Statistical Parametric Mapping 8 (SPM8) software (Wellcome Trust Centre for Neuroimaging, UCL Institute of Neurology, London) running on MathWorks Matrix Laboratory version 7 (MATLAB) (MathWorks Inc., Natick, Massachusetts). Preprocessing involved within-subject realignment, coregistration of the functional images to the individual three-dimensional structural scans, spatial normalization of the structural scans to the MNI template using unified segmentation, linear trend removal by high-pass filtering of the data with a cutoff of 128 seconds and spatial smoothing using a 8 mm Gaussian kernel.

Analysis at the first level was conducted using a single General Linear Model (GLM) for each individual in which each scan was coded as belonging to one of the nine experimental conditions. The 9 runs were modeled as one session within the design matrix, and nine regressors were used to model run-specific variation; six columns of realignment parameters were included as covariates of no interest.

3. Results

3.1 Behavioral, auditory, and cognitive tests

All participants had normal hearing; the mean pure-tone hearing threshold at 500, 1000, 2000, and 4000 kHz, averaged over both ears, was 5.4 dB HL (SE = 1.2 dB).

The mean SRT was -5.9 dB SNR (SE = .2 dB SNR, range -7.2 to -4.6). The mean TRT was 54.9% unmasked text, SE = .8%, range 51-61%, median = 53.4%. The mean RSpan was 25.3, SE = .8, range 20-31, median = 25. The TRT and RSpan performances are similar to those observed in other studies that examined comparable groups of subjects (e.g., Zekveld et al., 2007, 2011a).

3.2 Imaging session – behavioural data

Figure 2 shows the means and 95% within-subjects confidence interval (Loftus & Masson, 1994) of the word scores (percentage correct) in each of the six conditions in which sentences were presented.
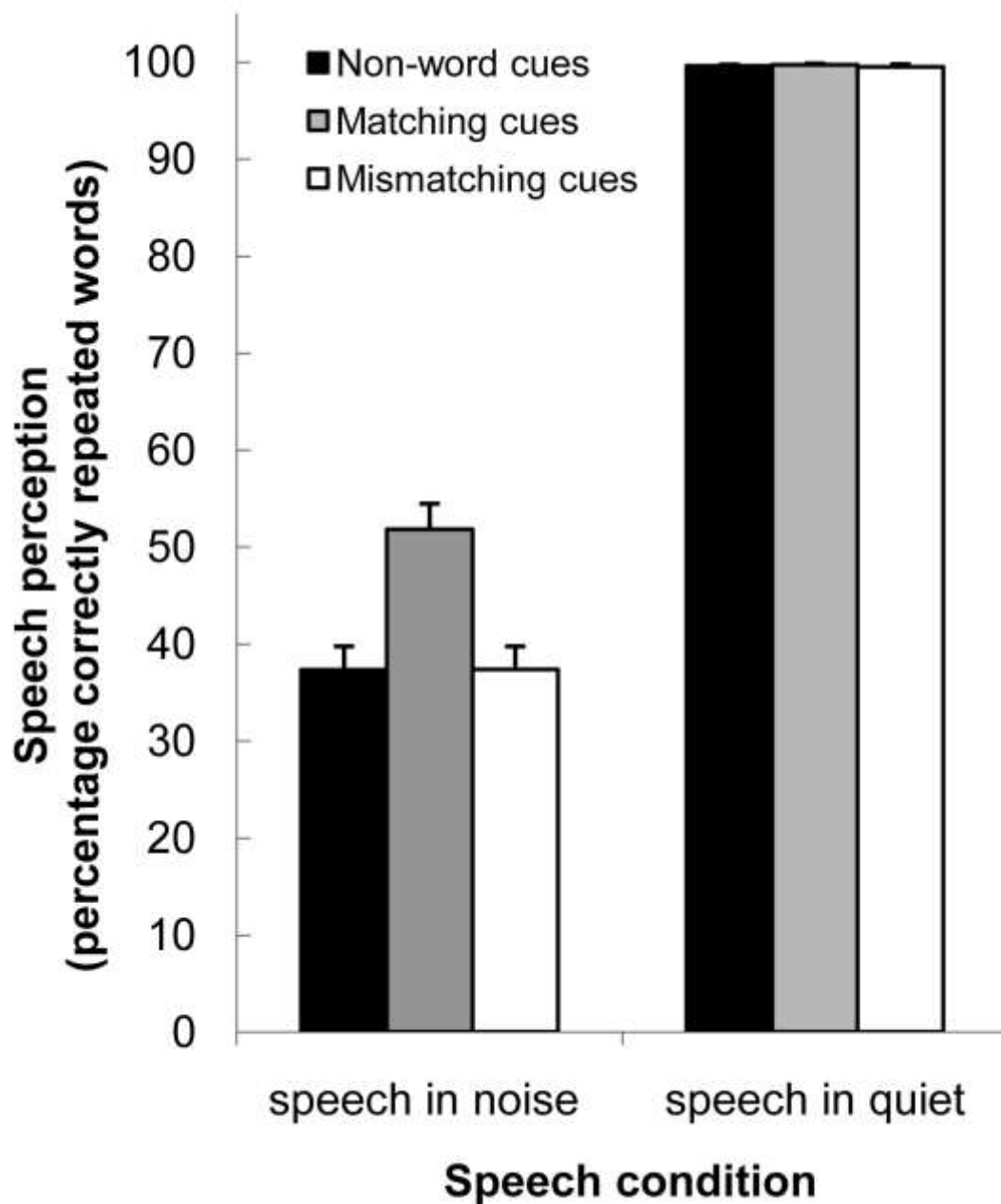

*Figure 2.*Means and 95% within-subjects confidence interval (Loftus & Masson, 1994; error bars) of the word scores (percentage of words that were correctly reported) in each of the 6 conditions of the cued speech-perception test in which sentences were presented.

We used repeated-measures analyses of variance (ANOVA) to test the main effects of *noise-level* (SIQ versus SIN) and *cue-type* (related cues, unrelated cues, or nonword baseline) and their interaction on the word-report scores. The ANOVA indicated main effects of *noise level* and *cue-type* ($F_{(1,17)} = 925$; $p < .001$ and $F_{(2,34)} = 18.2$, $\varepsilon = .9$; Greenhouse-Geisser corrected $p < .001$, respectively), as well as an interaction between noise level and cue-type ($F_{(2,34)} = 18.2$; $\varepsilon = .9$; Greenhouse-Geisser corrected $p < .001$). Figure 2 indicates that this interaction was based on greater intelligibility with related cues relative to both nonword and unrelated cues in the SIN conditions, and ceiling effects in the SIQ conditions. Differences among SIN conditions were confirmed by post-hoc testing. Word-report scores in the related SIN condition were significantly higher than the scores in the nonword baseline condition ($t_{(17)} = 4.6$, Bonferroni-corrected $p < .001$) and in the unrelated cue SIN condition ($t_{(17)} = 5.4$; Bonferroni-corrected $p < .001$).

3.3 Correlation analysis

Spearman correlation coefficients were calculated to examine the relation between word-report scores in the SIN conditions and the TRT and RSpan scores. We did not assess the associations with word report for the SIQ stimuli due to the ceiling performance in those conditions. We report non-parametric Spearman correlation coefficients as the sample size was relatively small (n = 18).

Regardless of cue type, TRT and RSpan were not statistically significantly ($p > .05$) associated with the percentage of words perceived correctly in the SIN conditions.

To examine the facilitatory or inhibitory effects of the cues, difference scores were calculated by subtracting the word report scores for the nonword cue conditions from word report scores for each of the related and unrelated cue conditions, within subjects, separately for each noise level.

In the SIN conditions, higher RSpan scores were associated with larger benefit obtained from the related cues (Spearman $r = .48$, $p < .05$; see Figure 3). The interfering effect of unrelated cues (i.e., the difference in word-scores between the unrelated and non-word cue conditions) was, on average, relatively small (see Figure 2). However, poorer TRT thresholds were related to more interference from the unrelated cues in the SIN condition (Spearman $r = -.49$, $p < .05$; see Figure 4). We did not observe a statistically significant association between the TRT and RSpan scores (Spearman $r = -.33$; $p = .18$).
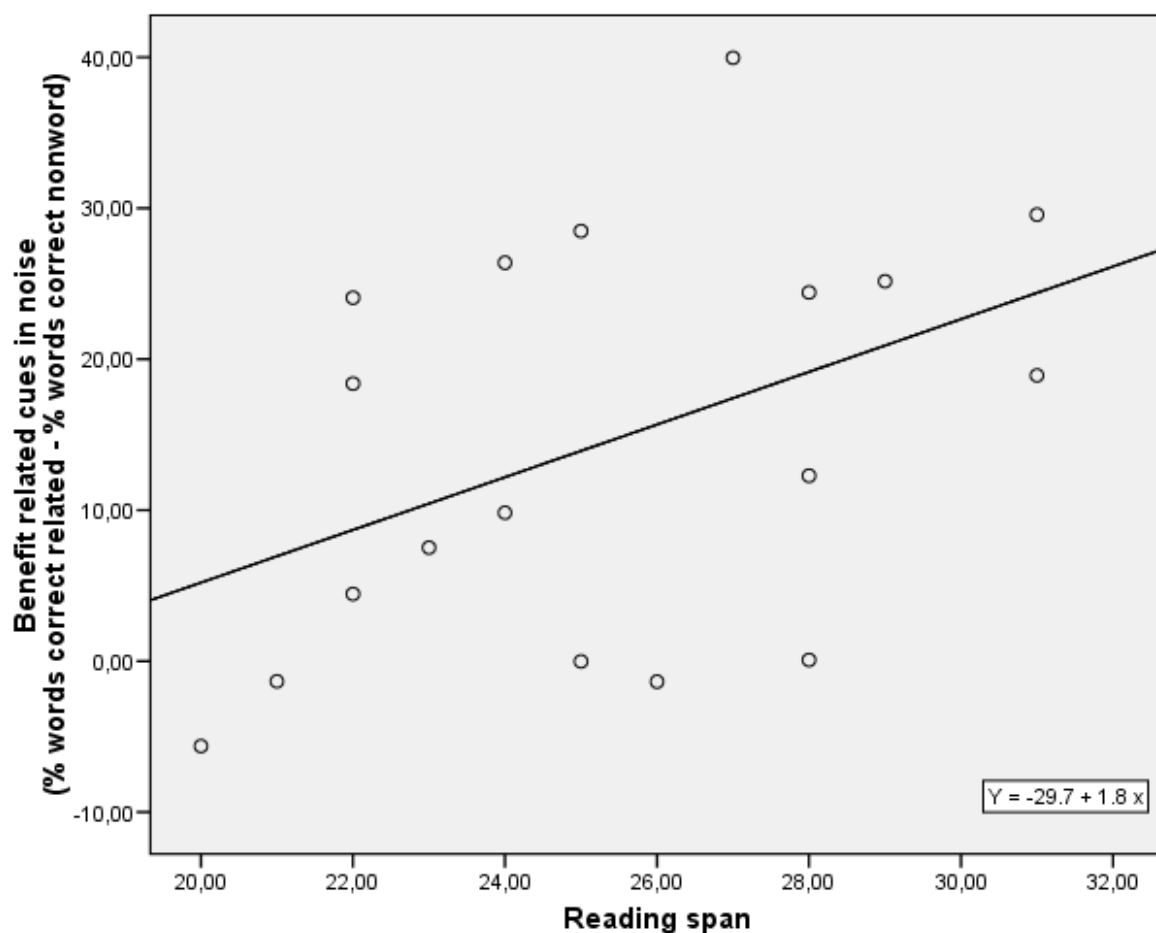

Figure 3. Relation between RSpan and benefit obtained from related cues for speech in noise stimuli (cue benefit = percentage of words correctly repeated in the related-cues condition minus the percentage of words correct in the nonword cue condition). Larger RSpan scores indicate larger working memory capacity.
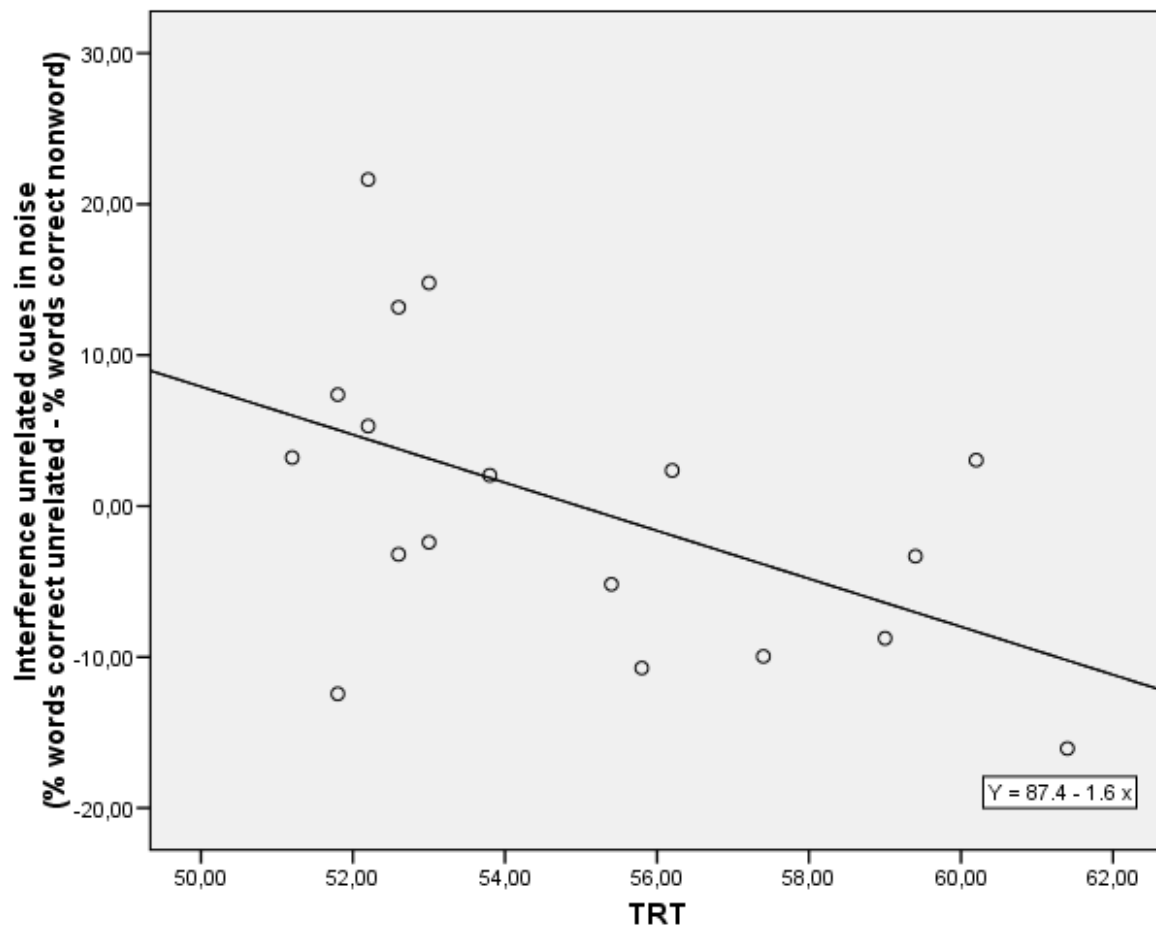
Figure 4. Relation between TRT and the interference from unrelated cues for speech in noise stimuli (cue interference = percentage of words correctly repeated in the unrelated-cues condition minus the percentage of words correct in the nonword cue condition). Lower TRTs indicate better performance.

3.4 fMRI results

Activity was considered statistically significant if it exceeded a voxel-based significance threshold of $p < .05$, corrected for multiple comparisons across the whole brain using the familywise error (FWE) correction based on Gaussian theory of random fields (Worsley et al., 1996). To check the quality of the functional MRI data, we first examined the effect of speech by using a t-test comparing the activation evoked by the speech in quiet and speech in noise conditions (with nonword cues) to that evoked by the noise alone condition with nonword cues (i.e., (SIQ + SIN)/2 > noise alone). Compared to noise alone, speech in noise and speech in quiet together evoked statistically significant activity in the STG bilaterally and in the left MTG (see Figure 5a). This has been observed in other studies as

well (e.g., Binder et al., 2004; Davis & Johnsrude, 2003; Obleser et al., 2007; Zekveld et al.,
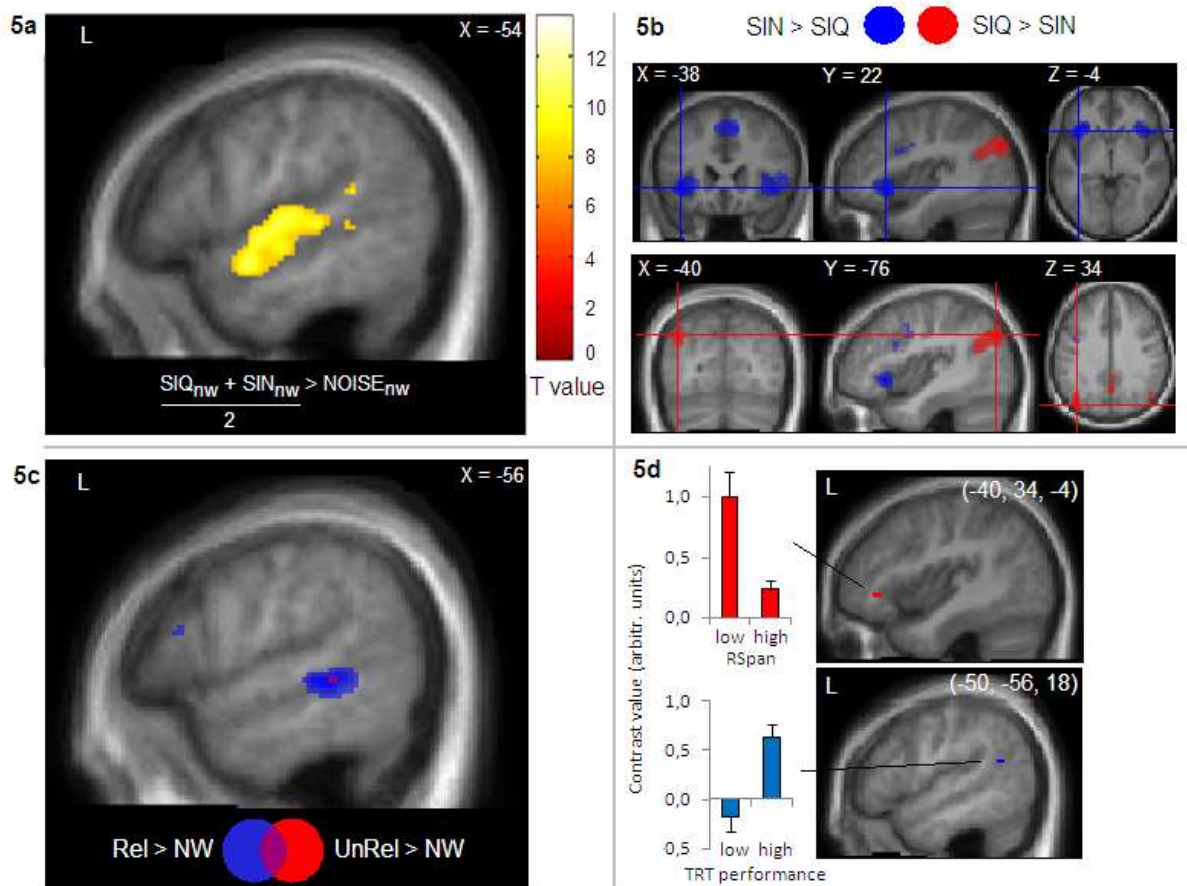
2006).

Figure 5a. Areas in the left superior and middle temporal gyrus and right superior temporal gyrus activated by both sentences in quiet (SIQ) and in noise (SIN) preceded by nonword (nw) cues, relative to stationary noise preceded by nonword cues. *5b.* Areas in that were differentially activated for speech in quiet versus speech in noise. The lower panel shows the activation peak of the area responding more to speech in quiet relative to speech in noise (red regions) and the upper panel shows the activation peak of the area responding more to speech in noise relative to speech in quiet (blue areas). *5c.* Areas in the left inferior frontal gyrus, left middle frontal gyrus, and left middle temporal gyrus were activated for sentences preceded by related (Rel) versus nonword (NW) cues (blue areas) and for sentences preceded by unrelated (UnRel) versus non-word cues (red areas). The activation evoked by the related relative to the nonword cues overlapped with the activation evoked by the unrelated cues in the middle temporal gyrus. 5d. Effects of cognitive subgroups (n=9) on the mean activation evoked by the six speech in noise and speech in quiet conditions relative to quiet. Subgroups were either based on the reading span (RSpan) or text reception threshold (TRT) performance. Error bars indicate standard errors. All activations are statistically significant at an FWE corrected threshold $p < .05$. The activation maps are superimposed on the mean structural scan in MNI space.

We examined the main effects of cue type (related, unrelated, nonword cues) and noise level (SIQ versus SIN) and their interaction on brain activation by entering subject-specific contrasts of the six conditions (compared to silence) into a second-level ANOVA, treating subjects as a random effect. The analysis indicated a significant main effect of noise level. Separate t-tests were performed to assess the origin of the main effect (see Figure 5b and Table 2). The results indicated that compared to SIQ, SIN evoked greater activity in an extensive network of regions including the superior frontal gyrus (bilaterally), the left precentral gyrus, left lateral orbitofrontal region, the left middle frontal gyrus, the right insular cortex and right IFG. Relative to SIN, SIQ evoked activation in the angular gyrus, and hippocampus, bilaterally; in the superior parietal region and parahippocampal gyrus in the left hemisphere, and in the lingual gyrus, precuneus and superior frontal gyrus in the right hemisphere.

The ANOVA additionally revealed a significant main effect of cue type in the left middle temporal gyrus, left inferior frontal gyrus and left supramarginal gyrus (see Table 3 and Figure 5c). To examine the origin of this main effect, we applied separate t-tests to compare cue types pairwise, collapsed across noise level (SIN and SIQ, see Table 3).

Compared to the non-word cue condition, the related-cue condition evoked more activation in the left MTG, left IFG and the left MFG (see Table 3 for coordinates), and the unrelated-cue condition elicited more activation in the left MTG and left IFG. The activation in the related-cue condition did not differ significantly from activation in the unrelated-cue condition.

In contrast to the behavioral results, The ANOVA did not indicate a statistically significant interaction between cue type and noise level, even with a fairly liberal statistical threshold (uncorrected $p < .05$, minimum cluster size = 15) and increased spatial smoothing (up to 12 mm Gaussian kernel). Because we hypothesized that the effect of cue on frontal-

Table 2. Significant (p < .05; FWE corrected) activation foci for areas sensitive to noise level.

Main effect of noise level

region	x	y	z	F
L SFG	-6	12	60	270
	-4	14	52	257
	-6	20	46	224
	0	50	-12	43.3
L OFG	-32	26	-2	197
L lat OFG	32	28	-2	188
L precentral gyrus	-44	2	44	123
	-56	-4	12	29.4
L IFG	-48	12	28	94.6
R MFG	48	42	26	67.5
	26	28	42	33.7
	28	4	54	31.3
R IFG	50	14	28	66
L MOG	-38	-76	34	114
L angular gyrus	-46	-58	25	85.8
L precuneus	-10	-60	24	93.9
R precuneus	4	-60	30	91.6
L cingulate gyrus	-6	-32	42	30.3
R angular gyrus	44	-62	24	62.8
L MFG	-36	22	50	60.9
	-22	26	42	36.3
	-32	52	18	27.3
L SFG	-24	32	54	28.8
L caudate	-14	14	4	54.8
R SFG	6	56	-8	49.5
	22	36	56	32.7
Left HC	-24	-6	-18	48.7
R precentral gyrus	62	-2	14	46.8
R supramarginal gyrus	54	-28	26	46.5
	44	-38	44	37.3
R putamen	28	-4	-10	42.9
	28	-10	2	42.6
L para HC gyrus	-30	-40	-12	38.1
L SPG	-40	-44	44	32.0
L MTG	-44	6	-36	31.9
R MTG	54	8	-36	30.8
L putamen	-32	-14	2	30.6
R caudate	14	12	2	28.3
L supramarginal gyrus	-62	-30	24	27.3

SIN > SIQ

region	x	y	z	T
L lat OFG	-38	22	-4	19.9
L precentral gyrus	-50	0	48	10.3
	-44	2	36	10.2
L SFG	0	16	54	14.8
	-4	10	66	14.1
R SFG	8	22	46	14.5
	10	32	26	10.7
R insular cortex	42	20	-6	14.1
	36	24	0	12.4
R IFG	50	20	4	13.9
	54	26	24	8.4
	48	14	28	7.9
	2	34	40	7.7
L MFG	-38	28	26	8.2
	-46	24	30	7.5

SIQ > SIN

region	x	y	z	T
L angular gyrus	-40	-76	34	18.0
	-44	-62	26	12.2
L SPG	-10	-62	22	11.9
R lingual gyrus	10	-54	8	11.4
R precuneus	2	-58	34	10.9
R angular gyrus	48	-64	28	10.9
	60	-58	20	8.4
L para HC gyrus	-26	-40	-14	10.2
R SFG	4	60	-8	10.0
	14	34	54	8.9
R HC	26	-10	-14	9.6
L HC	-26	-6	-20	9.5

L = left; R = right; SFG = superior frontal gyrus; OFG = orbito-frontal gyrus; IFG = inferior frontal gyrus; MOG -= middle occipital gyrus; MFG = middle frontal gyrus; HC = hippocampus; SPG = superior parietal gyrus; MTG = middle temporal gyrus

Table 3. Significant (p < .05; FWE corrected) activation foci (Brodmann area) for areas

sensitive to the cue type, collapsed over SIQ and SIN conditions.

Region	MNI peak coordinates			Z score	FWE corrected P
	x	*y*	*z*		
Main effect of cue					
L IFG (48)	-50	22	16	5.89	< .001
L IFG (44)	-52	22	26	5.10	.01
L MTG (21)	-52	-38	0	7.22	<.001
L supramarginal gyrus (40)	-60	-36	34	4.92	.023
Related > Nonword					
L IFG (37)	-44	42	-8	5.53	.004
L IFG (48)	-54	22	20	5.10	.025
L MFG (9)	-38	12	50	5.00	.038
L MTG (21)	-58	-34	0	6.75	< .001
Unrelated >Nonword					
L IFG (45)	-50	24	12	4.98	.042
L MTG (21)	-56	-36	0	4.98	.042

L = Left, R = Right, IFG = inferior frontal gyrus, MFG = middle frontal gyrus, MTG =

middle temporal gyrus.

lobe activity in particular would differ between SIQ and SIN conditions (see Table 4), we

examined the effect of cues separately on SIQ and SIN. For SIQ, the ANOVA indicated a

main effect of cue type in the left MTG and left IFG. Separate t-tests showed that the

Table 4. Significant ($p < .05$; FWE corrected) activation foci (Brodmann area) for areas sensitive to the cue types, separately for the speech in quiet (SIQ) and speech in noise (SIN) conditions. Contrasts between the cues that are not listed were not statistically significant.

Region	MNI peak coordinates			Z score	FWE corrected P
	x	*y*	*z*		
Main effect of cue					
SIQ L MTG (21)	-54	-38	0	5.94	< .001
L IFG (45)	-54	20	10	5.07	.011
SIN L MTG (21)	-50	-40	0	5.16	.007
Related > Nonword					
SIQ L STG (22)	-60	-36	2	5.60	< .001
SIN L MTG (21)	-60	-34	-4	6.16	< .001
Unrelated > Nonword					
SIQ n.s.					
SIN L MTG (21)	-50	-38	-2	5.11	.027

L = Left, R = Right, IFG = inferior frontal gyrus, MTG = middle temporal gyrus, STG = superior temporal gyrus, n.s. = not statistically significant

activation in the left STG was larger for the related compared to nonword cues. For SIN, we observed a main effect of cue in the left MTG only. Separate t-tests indicated that this main effect was based on larger activation evoked by related and unrelated compared to nonword cues. Thus, this analysis of simple effects confirmed that the main effect is observed at both noise levels (SIN and SIQ). These results suggest that the interaction observed behaviorally is

driven only by performance being at ceiling in the SIQ condition, and that the effect of cue type does not appear to depend on noise level, in terms of the neural processes recruited to support comprehension. Alternatively, BOLD response may not differentiate underlying processes in SIN and SIQ conditions.

Finally, we examined the effect of cognitive ability on activity levels by introducing individual RSpan and TRT scores as regressors in the 2nd level ANOVA. The abilities did not significantly interact with the main effects of noise level or cue-type on brain activation. As an alternative, less precise but more powerful approach, the same 2 x 3 full factorial model was analyzed in two additional second-level random effects ANOVAs, but with one added factor "cognitive group", resulting in a 2 x 2 x 3 factorial model. Two groups for each test (low and high RSpan and low and high TRT ability; subgroup n = 9) were formed by applying a median split on the RSpan and TRT scores. This approach has been adopted by others (e.g. Rudner et al., 2011) to examine the effects of cognition on listening under challenging conditions and allowed us to test the effect of cognitive group within the ANOVA model. The subgroup of 9 subjects with relatively poor TRT performance had a mean TRT of 57.6%, SE = .8%, range 53.8-61.4%, and the subgroup with relatively good TRT performance had a mean TRT of 52.3%, SE = .2%, range 51.2-53%. The mean RSpan of the subgroup of 9 subjects with relatively poor RSpan was 22.6, SE = .5, range 20-25, and the subgroup with relatively high RSpan had a mean RSpan of 28.1, SE = .7, range 25-31. For both the TRT and RSpan subgroups, mean SRTs did not differ statistically significantly between subgroups ($p >$ .42).

We examined the main effect of group and the interactions between group, noise level, and cue type on the brain response. For the average activation in the six SIQ and SIN conditions (compared to silence), we observed a main effect of RSpan group in the left IFG and left STG. A subsequent t-test on the average activation in the six conditions indicated

greater activity in these areas in the low RSpan subjects compared to the high RSpan subjects (IFG peak coordinates -40, 34, -4, $t = 5.7$, $p < .01$, see Figure 5d; STG peak coordinates -68, 2, -8, $t = 5.5$, $p < .05$ and -60, -12, -6, $t = 5.15$, $p < .05$).. There were no areas that were significantly more active in the high RSpan group compared to the low RSpan group.

We additionally observed an interaction effect between group and cue type. Activation evoked in the right middle frontal gyrus (peak coordinates 12, 36, 2, $t = 5.44$, $p < .05$) by sentences preceded by related vs. unrelated cues was greater in the low RSpan subjects compared to the high RSpan subjects.

The main effect of TRT group was not statistically significant. As the subgroups were relatively small (n = 9), we examined the directional t-tests comparing the average activation in the six SIQ and SIN conditions (compared to silence) between the TRT groups. The t-test comparing high TRT performers to low TRT performers revealed a significant difference in the left AG (peak coordinates -50, -56, 18, $t = 5.12$, $p < .05$, see Figure 5d). No areas were significantly more activated in the low TRT compared to the high TRT performers. There were no significant interaction effects between TRT group, noise level and / or cue type.

Regarding the association between the TRT, RSpan and the average brain activation evoked by the SIN and SIQ conditions (compared to silence), note that the subgroups were small and that the areas in which the activation differed, although statistically significant (FWE corrected p < .05), were relatively small as well. When applying a less stringent statistical threshold (p uncorrected < .0001), additional areas in the left and right STG were observed to be more active in the low RSpan subjects compared to the high RSpan subjects. In addition, in the most able subgroup of participants on the TRT task, the left MTG and left STG were more activated at this more liberal statistical threshold compared to less able participants.

4. Discussion

Both the behavioral and the fMRI data indicated an association between cued speech perception and interindividual differences in working memory capacity (reading span) and the ability to read masked written text (TRT test).

Reading span, which is considered a measure of working memory capacity (Daneman & Carpenter, 1980) was associated with the amount of benefit obtained from the related cues. The activation evoked by sentences in the left STG and left anterior IFG was larger for listeners with lower WM capacity. Left IFG activation was also modulated by cue type, with larger activation for sentences preceded by both related and unrelated cues relative to nonword cues. We additionally observed a larger difference in activity between related and unrelated cues (related > unrelated) in right medial frontal cortex for participants with higher as opposed to lower WM capacity. The behavioral and fMRI results hence both indicate that processes associated with using relevant context during speech perception are both more efficient and more successful in listeners with a high RSpan capacity. These processes could include the encoding of the cues in memory and evaluation of the relation between the cues and the speech content.

In general, activation in the angular gyrus was greater in individuals who were good at reading masked text than in those who were not. The behavioral data also indicated that this skill was associated with better ability to ignore the unrelated cues. This result is consistent with those observed in our previous behavioral study (Zekveld et al., 2011a). Humphries, Binder, Medler, and Liebenthal (2006) suggested that the AG is especially sensitive to stimuli that have combinatorial semantic structure, for example, when individual words fit together to form a complex meaning. The TRT test requires the reader to construct a meaningful sentence based on word-fragments. This involves combinatorial semantic processing as individual words are fitted together to produce a global semantic interpretation of the sentence.

Similar to the current data, Zekveld et al. (2011b) showed that better ability to read masked text was related to more cognitive processing load during sentence perception in noise as assessed by examination of the pupil response. These and the current data may suggest that participants with high TRT ability may use a particular strategy in both the TRT and speech perception tests that, although effortful, increases their performance over that observed in low-TRT individuals. This strategy may include active guessing operations on the content of the words in the unfolding sentence.

In the present study, three-word text cues presented prior to auditorily presented sentences in noise improved sentence perception if they were semantically related to the speech content (see also Zekveld et al., 2011a). Regardless of noise level, sentences preceded by both related and unrelated cues evoked more activation in the left MTG than did sentences preceded by non-word cues (see Tables 3 and 4). Activation in the left supramarginal gyrus and left IFG was also influenced by cue type, with more activation for both related and unrelated cues relative to nonword cues in left IFG (see Table 3). Additionally, the activation evoked in the left MFG was larger for related compared to unrelated cues. In the supramarginal gyrus, we observed a significant main effect of cue type, but no significant pairwise differences. The differential activation in response to meaningful cues compared to nonword cues in MTG, IFG, and MFG may reflect processes that differ between the processing of real words and nonwords, like visual analysis of word forms, and accessing phonological word forms and semantic representations (Joubert et al., 2004). It may also reflect differences in the interaction of the cues and the sentence between the meaningful and nonword cues, like processes associated with semantic retrieval, syntactical processing and controlled executive functions such as selection and evaluation (Kotz et al., 2002; Poldrack al., 1999; Just, Carpenter, Keller, Eddy, & Thulborn, 1996; Dapretto & Bookheimer, 1999; Humphries, Binder, Medler, & Liebenthal, 2007; Obleser et al., 2007).

In line with previous reports, noise level modulated activation in an extensive bilateral fronto-temporal network (see e.g., studies by Binder et al., 2000; Davis & Johnsrude, 2003; Friederici et al., 2010; Obleser & Kotz, 2010, and PET data reported by Mummery et al., 1999; see Table 2). Interestingly, the effect of cue type on patterns of activation did not appear to depend on noise level, in contrast to the word report scores (see Figure 2). Wild et al. (2012) performed an fMRI study in which they presented either an accurate sentence transcription, or random strings of consonants, simultaneously with a spoken sentence. Sentences were presented clearly, degraded but potentially intelligible, or completely unintelligible (signal-correlated noise). Several brain regions in the left temporal lobe were more activated by sentences accompanied by matching text compared to sentences accompanied by random letter strings. In contrast to the current findings, Wild et al. (2012) additionally observed an interaction between the type of text stimulus and speech type, such that the effect of text (matching vs letter strings) was greater in several temporal and frontal regions for degraded sentences relative to clear or unintelligible sentences. Obleser and Kotz (2010) also observed an interaction between the quality of the signal and the effect of context in their study of high- and low-cloze probability sentences (Gunter, Friederici & Schriefers, 2000). However, in their study, both the speech to be understood (the sentence-final word) and the context were equally degraded, and so the interaction they observed may reflect a decreased utility of context when signal quality was poor, and not a change in how context is used to support comprehension of degraded speech. Here, we avoided that potential confound by ensuring that the context was never degraded. Since we did not observe such an interaction in fMRI data, we tentatively conclude that the cognitive processes recruited to make use of semantic context are similar regardless of signal quality and that the behavioural interaction was mainly caused by the ceiling effect in the speech in quiet conditions.

We did not observe an activation difference between related and unrelated cues. This is surprising, since the intelligibility difference between these two conditions exceeded 10% (see Fig 2), and may indicate similar processing of stimuli in these two conditions despite different success (as manifest in word-report scores).

The fact that we did observe associations between the two cognitive tests and the neural activation during cued speech perception is remarkable in light of the fact that we did not observe the expected interaction between cue type and intelligibility. The cognitive group effects may indicate that different individual strategies for coping with related/unrelated cues driven by differences in cognitive ability may obscure an overall interaction between cue type and intelligibility. Whereas higher ability to withstand masking on the TRT task was associated with *more* activation in the left AG, higher reading span was associated with *less* activation in the left IFG, as is predicted by the ELU framework (Rönnberg et al., 2003). Similar to the current data, Francis and Nusbaum (2009) observed a relation between large working memory capacity and the efficiency with which synthetic speech could be recognized. Neubauer and Fink (2009) have reviewed studies addressing the relation between cognitive abilities and resource allocation. The relationship observed seems to depend on task difficulty, strategy differences, and the brain area under study. The current data tentatively point to different types of processes relevant for performing the RSpan and TRT tasks that are both also associated with cued speech perception. In the RSpan task, subjects have to make rapid semantic judgments while storing information in memory. This test was associated with more benefit obtained from the related cues and less activation in the STG and IFG during cued sentence perception. The TRT test taps into the ability to integrate masked word fragments into a meaningful sentence, and lower (better) thresholds were related to better ability to ignore unrelated cues and more activation in the AG. Recently, Besser et al. (2012) concluded that the TRT test and the RSpan test measure essentially different nonauditory

components of speech understanding in noise. Their study suggested a nearly age-independent association between TRT and SRT, and additionally indicated that the association between RSpan and speech perception is more sensitive to age-related declines.

In contrast to the current findings, Wong, Uppunda, Parrish, and Dhar (2008) observed that the STG was more active during listening to speech in noise relative to listening to speech in quiet. However, in line with the relation between RSpan and the STG activation as observed in the current study, they suggested a role of the STG in a phonological working memory network. It would be interesting to further assess the role of this area in future studies addressing interindividual differences in the activation evoked by speech in challenging listening conditions.

The associations among the two cognitive tests, the behavioral results and brain activation were observed despite the individual adaptation of the SNR applied in the SIN conditions. Thus, even when uncued speech perception performance was equalized across subjects, individual differences in the abilities tapped by the TRT and RSpan tests played a role in cued speech perception. This may suggest that these functions relate to the ability to make use of semantic context as such, besides the known relevance in uncued speech perception.

The individually adapted SNR resulted in 29% sentence perception in the uncued SRT test. However, the overall performance on speech-in-noise trials in the imaging session was on average below 29% correct. This relatively low performance may reflect the fact that listeners had to wait approximately 5 s before they were allowed to repeat the sentence in the imaging session, but not in the SRT test. In this interval, the functional images were acquired and the resulting scanner noise may additionally have affected the internal rehearsal and activation of the response. We speculatively suggest that this delay may have also potentiated the association between speech perception and reading span, as it presumably increased the

memory demands imposed by the test. In our previous behavioural study (Zekveld et al.,2011a), we applied the same approach as the one adopted in the current study. In that study, sentence perception performance in the cued speech perception test was around the targeted performance levels for the non-word cue conditions. Therefore, it is unlikely that the relatively low performances in the fMRI scanner observed here were caused by low test-retest reliability of the SRT procedure.

5. Conclusions

We manipulated availability of semantic context independently of the intelligibility of spoken sentences for the first time in a neuroimaging study. Semantic context was in the form of text cues that were presented prior to each of the spoken sentences. The cues were words that were related or unrelated to the spoken sentence, or non-words (neutral). Compared to the nonword baseline, sentences preceded by words evoked activation in the left IFG and MTG, irrespective of whether the words were semantically related to the sentences or not. This could relate to differences in the processing of real words versus nonwords, and/or may suggest a role for these areas in evaluating semantic context during speech comprehension. As predicted by the 'neural efficiency hypothesis' (Neubauer & Fink, 2009) and the ELU framework, higher working memory capacity was associated with better speech comprehension in noise with related cues, and less activation in the left anterior IFG and the left STG. The ability to read partly masked text was associated with less interference from the unrelated cues and increased activation of the left AG. This may support a role of this area in "combinatorial processing of semantic information", as suggested by Humphries et al. (2006). Overall, the present results suggest a left lateralized network for evaluating semantic context during speech comprehension that includes classical language processing areas and that is differentially modulated by distinct cognitive skills.

Acknowledgments

References

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

Andersson, U., Lyxell, B., Rönnberg, J., & Spens, K-E. (2001). Cognitive correlates of visual speech understanding in hearing-impaired individuals. *J Deaf Stud Deaf Educ, 6,* 103-116.

Arlinger, S., Lunner, T., Lyxell, B., & Pichora-Fuller, K. (2009).The emergence of cognitive hearing science. *Scand J Psychol, 50, 371-384.*

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995).The CELEX Lexical Database (Version Release 2) [CD-ROM]. Philadelphia: Linguistic Data Consortium, University of Pennsylvania.

Baddeley, A., Logie, R., Nimmo-Smith, I., & Brereton, N. (1985) Components of fluent reading. *J Mem Lang, 24,* 119–131.

Bailey, I. L., & Lovie, J. E. (1980).The design and use of a near-vision chart. *Am J Optom Physiol Opt, 57,* 378-387.

Besser, J., Zekveld, A. A., Kramer, S. E., Rönnberg, J., & Festen, J. M. (2012). New measures of masked text recognition in relation to speech-in-noise perception and their associations with age and cognitive abilities. *(In press, J Speech Lang Hear Res).*

Binder , J. R. & Desai, R. H. (2011). The neurobiology of semantic memory. *Trends Cogn Neurosci, 15,* 527-536.

Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S. F., Springer, J. A., Kaufman, J. N., et al. (2000). Human temporal lobe activation by speech and nonspeech sounds. *Cereb Cortex, 10*, 512-528.

Binder, J. R., Liebenthal, E., Possing, E. T., Medler, D. A., & Douglas Ward, B., (2004). Neural correlates of sensory and decision processes in auditory object identification. *Nat Neurosci, 7*, 295-301.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *J Verb Learn Verb Behav, 19,* 450-466

Dapretto, M., & Bookheimer, S.Y. (1999). Form and content: dissociating syntax and semantic in sentence comprehension. *Neuron, 24,* 427-432.

Davis, M. H., & Johnsrude, I. S. (2003). Hierarchical processing in spoken language comprehension. *J Neurosci, 23*, 3423-3431.

Drager, K. D. R., & Reichle, J. E. (2001). Effects of discourse context on the intelligibility of synthesized speech for young adult and older adult listeners: Applications for AAC. *J Speech Lang Hear Res, 44,* 1052-1057.

Engle, R.W. (2002). Working memory capacity as executive attention. *Curr Dir Psychol Sci, 11,* 19-23.

Francis, A. L. & Nusbaum, H. C. (2009). Effects of intelligibility on working memory demand for speech perception. *Atten Percept Psychophys, 71,* 1360-1374.

Friederici, A. D., Kotz, S. A., Scott, S. K., & Obleser, J. (2010).Disentangling syntax and intelligibility in auditory language comprehension. *Hum Brain Mapp, 31*, 448-457.

Gunter, T. C., Friederici, A. D., & Schriefers, H. (2000). Syntactic gender and semantic expectancy: ERPs reveal early autonomy and late interaction. *J Cogn Neurosci, 12*, 556-568.

Hagoort, P., & van Berkum, J. (2007). Beyond the sentence given. *Philos Transact A Math Phys Eng Sci, 362*, 801-811.

Hall, D. A., Haggard, M. P., Akeroyd, M. A., Palmer, A. R., Summerfield, A. Q., Elliott, M. R., Gurney, E. M. et al., (1999). "Sparse" temporal sampling in auditory fMRI. *Hum Brain Mapp, 7,* 213-223.

Humphries, C., Binder, J. R., Medler, D. A., & Liebenthal, E. (2006).Syntactic and semantic modulation of neural activity during auditory sentence comprehension. *J Cogn Neurosci, 18,* 665-679.

Humphries, C., Binder, J. R., Medler, D. A., & Liebenthal, E. (2007). Time course of semantic processes during sentence comprehension: An fMRI study. *Neuroimage, 36,* 924-932.

Joubert, S., Beauregard, M., Walter, N., Bourgouin, P., Beaudoin, G., Leroux, J.-M., Karama, S., Lecours, A.R., 2004. Neural correlates of lexical and sublexical processes in reading. *Brain Lang, 89,* 9–20.

Just, M. A., Carpenter, P. A., Keller, T. A., Eddy, W. F., & Thulborn, K. R. (1996). Brain activation modulated by sentence comprehension. *Science, 4,* 114-116.

Just, M. A., Carpenter, P. A., & Miyake, A. (2003). Neuroindices of cognitive workload: Neuroimaging, pupillometric and event-related potential studies of brain work. *Theor Issues Ergonomics Sci, 4,* 56-88.

Kane, M. J., & Engle, R.W. (2000). Working-memory capacity, proactive interference, and divided attention: Limits on long-term memory retrieval. *J Exp Psychol Learn Mem Cogn, 26,* 336-358.

Kramer, S. E., Zekveld, A. A., & Houtgast, T. (2009). Measuring cognitive factors in speech comprehension: The value of using the text reception threshold test as a visual equivalent of the SRT test. *Scand J Psychol, 50,* 507-515.

Kotz, S. A., Cappa, S. F., von Cramon, D. Y., & Friederici, A. D. (2002). Modulation of the

lexical-semantic network by auditory semantic priming: An event-related functional

fMRI study. *Neuroimage, 17,* 1761-1772.

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *J Acoust Soc Am, 49*,

467-477.

Loftus, G. R., & Masson, M. E. (1994). Using confidence intervals in within-subject designs.

Psychon Bull Rev, 1, 476-490.

Mäntylä, T. & Nilsson, L-G. (1988). Cue distinctiveness and forgetting: Effectiveness of self-

generated retrieval cues in delayed recall. *J Exp Psychol: Learn Mem Cogn, 14,* 502-

509.

Mason, R. A., & Just, M. A. (2007). Lexical ambiguity in sentence comprehension. *Brain

Res, 1146,* 115-127.

Miller, G. A., Heise, G. A., & Lichten, W. (1951).The intelligibility of speech as a function of

the context of the test materials. *J Exp Psychol, 41,* 329-335.

Miller, G. A. & Isard, S. (1963). Some perceptual consequences of linguistic rules. *J Verb

Learn Verb Behav, 2,* 217-228. Moll., K., Cardillo, E., & Aydelott-Utman, J. (2001).

Effects of competing speech on sentence-word priming: Semantic, perceptual, and

attentional factors. In J. D. Moore & K. Stenning (Eds.), *Proceedings of the Twenty-

third Annual Conference of the Cognitive Science Society* (pp. 651–656). Mahwah, NJ:

Lawrence Erlbaum Associates, Inc.

Mummery, C. J., Ashburner, J., Scott, S. K., & Wise, R. J. S. (1999). Functional

neuroimaging of speech perception in six normal and two aphasic subjects. *J Acoust

Soc Am, 106,* 449-457.

Nakano, H., Saron, C., & Swaab, T. Y. (2010). Speech and span: Working memory capacity impacts the use of animacy but not of world knowledge during spoken sentence comprehension. *J Cogn Neurosci, 22,* 2886-2898.

Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychol Rev, 115,* 357-395.

Obleser, J., & Kotz, S. A. (2010). Expectancy constraints in degraded speech modulate the language comprehension network. *Cereb Cortex, 20*, 633-640.

Obleser, J., Wise, R. J. S., Dresner, & Scott, S. K. (2007). Functional integration across brain regions improves speech perception under adverse listening conditions. *J Neurosci, 27,* 2283-2289.

Osaka, M., Osaka, N., Kondo, H., Morishita, M., Fukuyama, H., Aso, T., et al., (2003). The neural basis of individual differences in working memory capacity: An fMRI study. *Neuroimage, 18,* 789-797.

Pichora-Fuller, M. K., Schneider, B. A., & Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *J Acoust Soc Am, 97,* 593-608.

Plomp, R. (2002). *The Intelligent Ear.* Mahwah, NJ: Lawrence Erlbaum Associates.

Plomp, R., & Mimpen, A. M. (1979). Improving the reliability of testing the speech reception threshold for sentences. *Audiology, 18,* 43-52.

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, 15-35.*

Rönnberg, J. (2003). Cognition in the hearing impaired and deaf as a bridge between signal and dialogue: a framework and a model. *Int J Audiol, 42,* S68-S76.

Rönnberg, J., Arlinger, S., Lyxell, B., & Kinnefors, C. (1989). Visual evoked potentials: Relation to adult speechreading and cognitive function. *J Speech Hear Res, 32,* 725-735.

Rönnberg, J., Rudner, M., Foo, C.,& Lunner, T. (2008). Cognition counts: A working memory system for ease of language understanding (ELU). *Int J Audiol, 47, S2,* S99-S105.

Rönnberg, J., Rudner, M., Lunner, T., & Zekveld, A. A. (2010). When cognition kicks in: Working memory and speech understanding in noise. *Noise Health, 12,* 263-269.

Rudner, M., Rönnberg. J. & Lunner, T. (2011). Working memory supports listening in noise for persons with hearing impairment. *J Am Acad Audiol, 22*, 156-167.

Sörqvist P., Ljungberg, J. K., & Ljung R. (2010). A sub-process view of working memory capacity: evidence from effects of speech on prosememory. *Memory 18*, 310–326.

Tesink, C. M. J. Y., Magnus Petersson, K. M., Van Berkum, J. J. A., Van den Brink, D., Buitelaar, J. K., & Hagoort, P. (2009). Unification of speaker and meaning in language comprehension: An fMRI study. *J Cogn Neurosci, 21,* 2085-2099.

Van Strien, J.W. (1992). Classificatie van links- en rechtshandige proefpersonen. *Ned Tijdschr Psychol. 47*, 21.

Versfeld, N. J., Daalder, L., Festen, J. M., & Houtgast, T., (2000). Method for the selection of sentence materials for efficient measurement of the speech reception threshold. *J Acoust Soc Am, 107,* 1671-1684.

Wingfield, A. & Tun, P. A. (2007). Cognitive supports and cognitive constraints on comprehension of spoken language. *J Am Acad Audiol, 18,* 548-558.

Wild, C. J., Davis, M. H., & Johnsrude, I. S. (2012). Human auditory cortex is sensitive to the perceived clarity of speech. *NeuroImage,* http://dx.doi.org/10.1016/j.neuroimage.2012.01.035

Wong, P. C. M., Uppunda, A. K., Parrish, T. B., & Dhar, S. (2008). Cortical mechanisms of speech perception in noise. *J Speech Lang Hear Res, 51, 1026-1041.*

Worsley, K. J., Marrett, S., Neelin, P., Vandal, A. C., Friston, K. J., & Evans, A. C. (1996). A unified statistical approach for determining significant voxels in images of cerebral activation. *Hum Brain Mapp, 4,* 58-73.

Zekveld, A. A., Heslenfeld, D. J., Festen, J. M., & Schoonhoven, R. (2006). Top-down and bottom-up processes in speech comprehension. *Neuroimage, 32*, 1826-1836.

Zekveld, A. A., George, E. L. J., Kramer, S. E., Goverts, S. T. & Houtgast, T. (2007). The development of the Text Reception Threshold test: A visual analogue of the Speech Reception threshold test. *J Speech Lang Hear Res, 50,* 576-584.

Zekveld, A. A., Kramer, S. E., & Festen, J. M. (2011b).Cognitive load during speech perception in noise: The influence of age, hearing loss, and cognition on the pupil response. *Ear Hear, 32,* 498-510

Zekveld, A. A., Rudner, M., Johnsrude, I. S., Festen, J. M., Van Beek, J. H. M., & Rönnberg, J. (2011a). The influence of semantically related and unrelated text cues on the intelligibility of sentences in noise. *Ear Hear, 32,* e16-e25.