The Interface Between Language and Attention: Prosodic Focus Marking Recruits a General Attention Network in Spoken Language Comprehension

Line Burholt Kristensen[1], Lin Wang[2,3,4], Karl Magnus Petersson[2,3] and Peter Hagoort[2,3]

[1]Department of Scandinavian Studies and Linguistics, University of Copenhagen, Copenhagen, Denmark [2]Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands [3]Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands and [4]Key Laboratory of Behavioral Science, Institute of Psychology, Chinese Academy of Sciences, Beijing, China

Address correspondence to P. Hagoort, Max Planck Institute for Psycholinguistics, PO Box 310, 6500 AH Nijmegen, The Netherlands. Email: peter.hagoort@mpi.nl

L.B.K. and L.W. contributed equally to the work, and the order is alphabetical.

In spoken language, pitch accent can mark certain information as focus, whereby more attentional resources are allocated to the focused information. Using functional magnetic resonance imaging, this study examined whether pitch accent, used for marking focus, recruited general attention networks during sentence comprehension. In a language task, we independently manipulated the prosody and semantic/pragmatic congruence of sentences. We found that semantic/pragmatic processing affected bilateral inferior and middle frontal gyrus. The prosody manipulation showed bilateral involvement of the superior/inferior parietal cortex, superior and middle temporal cortex, as well as inferior, middle, and posterior parts of the frontal cortex. We compared these regions with attention networks localized in an auditory spatial attention task. Both tasks activated bilateral superior/inferior parietal cortex, superior temporal cortex, and left precentral cortex. Furthermore, an interaction between prosody and congruence was observed in bilateral inferior parietal regions: for incongruent sentences, but not for congruent ones, there was a larger activation if the incongruent word carried a pitch accent, than if it did not. The common activations between the language task and the spatial attention task demonstrate that pitch accent activates a domain general attention network, which is sensitive to semantic/pragmatic aspects of language. Therefore, attention and language comprehension are highly interactive.

Keywords: attention network, FMRI, information structure, pitch accent, sentence processing

Introduction

In classical models of sentence comprehension of either the garden-path variety (Frazier and Clifton 1997) or in the constraint-based framework (Trueswell et al. 1994), the implicit assumption is usually that a full phrasal configuration results and a complete interpretation of the input string is achieved. However, often the listener interprets the input on the basis of bits and pieces that are only partially analyzed. As a consequence, the listener might overhear semantic information (Moses illusion, Erickson and Mattson 1981) or syntactic information (Chomsky illusion, Wang et al. in revision). Ferreira et al. (2002) introduced the phrase "good-enough processing" to refer to the listeners' and readers' interpretation strategies. Thus, in contrast to what classical models of sentence processing implicitly assume, the depth of processing varies across the language input. This might be the reason that language makes use of prosodic or syntactic devices to guide the depth of processing.

This aspect of linguistic meaning is known as "information structure" (IS; Halliday 1967; Büring 2007). The IS of an utterance essentially focuses the listener's attention on the crucial (new) information in it. In languages such as English and Dutch, prosody plays a crucial role in marking IS. For instance, the new or relevant information will typically be pitch accented. After a question like *What did Mary buy at the market?*, the answer might be *Mary bought VEGETABLES* (accented word in capitals). In this case, *vegetables* is the focus constituent, which corresponds to the information provided for the wh-element in the question.

There is no linguistic universal for signaling IS. The way IS expressed varies within and across languages. In some languages it may impose syntactic locations for the focus constituent, in other languages focus-marking particles are used, or prosodic features like phrasing and accentuation (Kotschi 2006; Miller et al. 2006; Gussenhoven 2008).

According to many linguistic theories, IS is an aspect of the core machinery of language and part of the syntactic or prosodic representation (Beckman 1996; Büring 2007). However, we investigate a new and alternative proposal, which does not reduce IS to a brick in either syntactic or prosodic representation. Our proposal is that languages provide linguistic means, such as prosodic or syntactic marking of new information (IS marking), that are triggers for recruiting a general attentional network in the brain in the service of increased processing of the marked constituents. Within a good-enough processing framework, this might be a safeguard against the possibility that the listener might overhear the most relevant bits and pieces of the linguistic input.

It has been shown that linguistically focused elements receive more attention than background information (Hornby 1974; Cutler and Fodor 1979; Birch and Rayner 1997). The strong bond between linguistic focus and attention is further supported by studies of how IS modulates the so-called semantic illusion effect. The term semantic illusion refers to an effect found in a study by Erickson and Mattson (1981), in which readers were presented with sentences containing a subtle world knowledge anomaly, e.g. *How many animals of each kind did Moses take on the Ark?* Forty-eight percent of readers gave the answer "two" without noticing that the sentence contained an anomalous word (*Moses*), as this word was semantically related to the correct word (*Noah*), thereby creating a semantic illusion. IS can modulate this semantic illusion effect. Subjects are more likely to notice anomalies when the anomalous word is marked as focus by an *it*-cleft

structure (*It was Moses who took 2 animals of each kind on the Ark.* vs. *It was 2 animals of each kind that Moses took on the Ark.* Bredart and Modolo 1988) or by capitalization (Bredart and Docquier 1989). These results suggest that focused elements are subject to more attention and more elaborate processing than non-focused elements. This claim is further supported by 2 event-related potential studies (Wang et al. 2009, 2011). In response to semantic anomalies, there was a larger N400 effect when the anomalous word was linguistically focused (by means of a wh-question context during reading: Wang et al. 2009; or by means of pitch accent during listening: Wang et al. 2011). These results are in line with the suggestion that IS directs the reader's attention toward focused constituents, leading to a more detailed processing of focused constituents. Although the association between IS and attention is extensively discussed, it remains an open question which attention networks are engaged in processing IS. It is not clear whether the brain network involved in this attention modulation constitutes a shared network with non-linguistic attention. The first question of our study is therefore: does the attention network modulated by IS constitute a separate network, or is it part of a shared general attention network?

Corbetta and Shulman (2002) identified 2 attention pathways: a dorsal fronto-parietal network and a ventral fronto-parietal network. The dorsal attention network includes the intraparietal sulcus (IPS) and superior parietal lobe, as well as the dorsal frontal cortex along the precentral sulcus, near or including the frontal eye field. It mediates the allocation of top–down attention driven by knowledge, expectations, or current goals. The ventral attention system involves the inferior parietal lobe (IPL) and the ventral frontal cortex, including parts of the middle frontal gyrus (MFG), the inferior frontal gyrus (IFG), and the anterior insula. It mediates bottom–up attention driven by relevant stimuli, especially unexpected and novel ones. Besides these 2 networks, subcortical structures such as superior colliculus and pulvinar nucleus of the thalamus are also important in coordinating attention (Shipp 2004).

The attention networks have primarily been investigated in the domain of spatial attention (Corbetta and Shulman 2002; Meyer et al. 2003; Vossel et al. 2006; Salmi et al. 2007). Several studies indicate that the brain regions engaged in spatial attention are also engaged in language-related processes. These processes include the maintenance of a linguistic focus in working memory (Osaka et al. 2007), directing attention toward semantic categories (Cristescu et al. 2006), as well as directing attention toward specific acoustic properties of speech sounds, such as attention to female versus male voices (Shomstein and Yantis 2006), to high versus low pitch (Hill and Miller 2010), to linguistic versus emotional aspects of intonation (Wildgruber et al. 2004), to lexical tones (Li et al. 2003, 2010), and to contrastive stress in sentences (Tong et al. 2005; Perrone et al. 2010). Although behavioral studies have established a link between attention and IS, the neuroimaging literature has not explored whether attention and IS are also related in terms of activated brain regions. The relation between attention and IS can shed light on whether IS processing should be viewed as modular (i.e. internal to the language system) or as relying on general attention mechanisms. On the grounds of activation of attention networks by language-related processes, we hypothesized that

pitch accent, as a linguistic marker of focus, will recruit general attention networks during language processing.

As mentioned earlier, IS can modulate semantic processing: focused information is more deeply processed. We therefore expect that the IS of stimuli will modulate unification processes, that is, operations (either phonological, syntactic, or semantic) that combine word information into larger units. The unification process for a sentence with a focused anomaly is assumed to be different from a sentence with an unfocused anomaly in terms of unification load. Generally, the unification load of anomalous sentences is higher than that of congruent sentences. This leads to increased BOLD-response in the unification area in the left IFG (LIFG: Hagoort 2005; Hagoort et al. 2009). This effect is seen for various kinds of semantic (Zhu et al. 2009) and pragmatic anomalies, including world knowledge anomalies (Hagoort et al. 2004; Menenti et al. 2009), mismatch between speaker and sentence content (Tesink et al. 2009), and mismatch between expected focus and placement of pitch accent (van Leeuwen et al. submitted). Therefore, the undetected violations might attenuate the LIFG activation. The present study examines whether the processing of IS marking, which is presumed to modulate attention networks, can further modulate brain regions that are known to be involved in language unification (i.e. the LIFG).

More specifically, we independently manipulated the prosody (one way to express IS) and the congruence of sentences to examine the interplay between the processing of IS and unification. We predicted that the processing of IS markers would activate attention networks, which would further modulate the unification process. To test whether IS markers activate a domain general attention network, we also performed an auditory spatial attention task. The identified attention network in this task was compared with IS-related activations in the language task, i.e. attention specifically engaged in unification processes during language comprehension. We predicted that we would find overlaps between these activations and the activations of the attention localizer, if a general attention network is recruited during language processing.

Methods

Participants

Twenty-four university students (mean age 21 years, range 18–24; 6 males) served as paid volunteers. All were right-handed native speakers of Dutch, with normal or corrected to normal vision. None of them had hearing problems, dyslexia, or a history of neurological or psychiatric diseases. Informed consent was obtained before the experiment. Thirteen additional subjects were scanned, but excluded from analysis because of technical problems with the magnetic resonance scanner (10 subjects) or poor task performance (3 subjects, see below).

Stimuli and Procedures

The experiment consisted of 2 tasks: an auditory spatial attention task (localizer task) with non-linguistic auditory stimuli (beeps) and a spoken language task with spoken language stimuli. Both tasks were auditory to avoid comparisons between different modalities. We chose a spatial attention task because the effects of spatial attention is well established compared with, for instance, feature-based attention like attention to form or color. The auditory spatial attention task was employed to localize the attention network in the particular group of participants, while the language task was used to investigate whether

IS (marked by pitch accent) activates parts of the attention network observed in the auditory spatial attention task, and whether this network modulates the process of language unification.

The Auditory Spatial Attention Task

Stimuli. Two tones of different frequency served as auditory stimuli: a cue (600 Hz) and a target (800 Hz) tone. Each tone lasted 150 ms, with a 10 ms linear onset–offset ramp. We presented the stimuli via earplugs, which also attenuated the scanner noise.

Procedure. The auditory spatial attention task was a modified version of the visual attention paradigm by Corbetta et al. (2000), which clearly distinguished 2 attention networks. There were 4 trial types: valid, invalid, noise, and cue-only trials (see Fig. 1). On valid cue trials (40% of the total trials), the subjects heard a cue tone in either the left or the right ear, and this cue tone indicated the location of the upcoming target tone (i.e. a left ear cue tone would predict a left ear target tone). After the presentation of the cue tone, the subject would experience a 1.5–3 s silent period followed by 8–15 binaurally presented filler tones (600 Hz, same frequency as the cue tone), which were then followed by the target. Finally, an end-of-trial white circle "O" was displayed at the center of the screen for 1.5–3 s. The invalid cue trials (20%) were similar to valid cue trials, except that the target appeared at the un-cued location (i.e. if the cue predicted a left ear location, the target would appear in the un-cued right ear). When the end-of-trial signal appeared for valid and invalid cue trials, the subjects were required to press a response button as quickly as possible upon detecting the target: a left button—"left side target"—to be pressed with the right index finger, and a right button—"right side target"—to be pressed with the right middle finger. The noise trials (20%) were similar to valid and invalid cue trials, except that the trial ended after the filler tone and no target tone was presented. Finally, on cue-only trials (20%), the subject heard a cue tone, but no filler or target tones appeared in either ear. The cue-only trials ended 1 s after cue presentation. No button presses were required for noise trials and cue-only trials, as these 2 types did not have target tones. Overall, the cue correctly predicted the target location on two-thirds
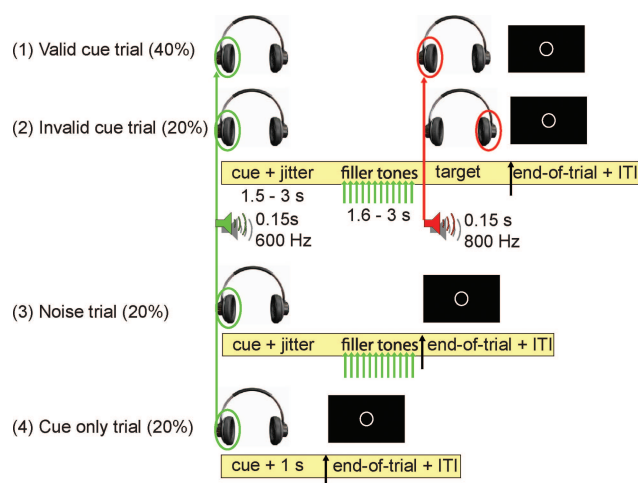
of trials in which a target was presented. The cue tone was used to orient attention toward left or right. The purpose of the filler tone period was for maintaining the top-down attention. The contrast between the target (including both valid and invalid targets) and noise trials was meant to reflect stimulus-driven bottom-up attention caused by the target tone, and the contrast between the invalid and valid cue trials indicated the re-orientation of attention to task-relevant stimuli.

In total, there were 150 trials. The 4 trial types were presented in a pseudorandom way, and no more than 3 trials with the same condition were presented in succession. A practice session containing 15 trials was conducted outside the scanner to familiarize each subject with the procedure. The subjects were instructed to maintain their attention on the cue location during the filler tone period (which lasted 1.6–3 s, depending on the number of filler tones). To minimize head movements, we asked the subjects to look at a white fixation cross displayed at the center of the screen throughout the task, except when the end-of-trial "O" appeared.

The Spoken Language Task

Stimuli. We constructed 200 Dutch sentence quartets. Within each quartet, we independently manipulated a specific word or noun phrase (defined as the Critical Word, CW) along 2 factors: Congruence (Congruent: C+; Incongruent: C−) and Prosodic pattern (With pitch accent: P+; Without pitch accent: P−). The congruence of the CW was manipulated so it either fitted the sentence context (C+) or violated semantics or general world knowledge (C−). The C− and C+ sentences were identical except for the CWs (see Table 1 for example). For an optimal semantic illusion effect, we constructed quartets for which the congruent and incongruent CWs were semantically related. We manipulated IS by either placing a pitch accent on the CW (P+) or not (P−).

The 200 quartets were recorded by a male native Dutch speaker at 44.1 kHz sampling rate and 16-bit resolution in a soundproof recording room. Praat 4.0 (Boersma and Weenink 2002) was used to normalize loudness differences between sentences by scaling the intensity to 70 dB. We conducted a pretest in order to ensure that the sentences met the following criteria: (1) the content of the incongruent versions should be incompatible with semantic/pragmatic knowledge of an average young native speaker of Dutch; (2) the content of the 2 congruent versions (C+) should be unproblematic, that is, the C+ sentences were taken as congruent no matter if the CW was realized with a pitch accent (P+) or not (P−).

Material Selection. The 200 quartets of sentences were arranged into 4 lists, with 200 items per list (50 items per condition) using a Latin square procedure. Therefore, exactly 1 exemplar of all the 200 quartets was presented in each list. The 4 lists were presented to 20 native Dutch speakers (mean age 21 years, range 19–24, 5 males)

Figure 1. The procedure of the auditory spatial attention task. (1) In the valid cue trial, a cue tone (600 Hz, 0.15 s, shown in green) is presented in 1 ear to indicate the location of the upcoming target tone. After a 1.5–3 s silent period (jitter), 8–15 filler tones (with the same frequency and duration as the cue tone) are presented binaurally (indicated by the green arrows). Then a target tone (800 Hz, 0.15 s, shown in red) is presented in the cued position. Finally, an end-of-trial white circle "O" is displayed at the center of the screen for 1.5–3 s (ITI). (2) The invalid cue trial is similar to the valid cue trial, except that the target appears at the un-cued location. (3) The noise trial is similar to the valid and invalid cue trials, except that the trial ends after the filler tones and no target tone is presented. (4) The cue only trial only contains a cue tone, and no filler or target tones are presented in either ear. The cue trial ends 1 s after the cue presentation.

Table 1

Exemplification of the 4 conditions of 1 item in the language task

(1) C+P+ (Congruent, With pitch accent)
Volgens het Bijbelboek Genesis bracht NOACH twee dieren van iedere soort op de ark. (According to the book of Genesis, NOAH brought 2 animals of each kind on the ark.)

(2) C−P+ (Incongruent, With pitch accent)
Volgens het Bijbelboek Genesis bracht MOZES twee dieren van iedere soort op de ark. (According to the book of Genesis, MOSES brought 2 animals of each kind on the ark.)

(3) C+P− (Congruent, Without pitch accent)
Volgens het Bijbelboek Genesis bracht Noach twee dieren van iedere soort op de ark. (According to the book of Genesis, Noah brought 2 animals of each kind on the ark.)

(4) C−P− (Incongruent, Without pitch accent)
Volgens het Bijbelboek Genesis bracht Mozes twee dieren van iedere soort op de ark. (According to the book of Genesis, Moses brought 2 animals of each kind on the ark.)

Note: The examples were originally in Dutch. Literal translations in English are given in brackets. The critical words (CWs) are underlined, and the words with pitch accent are in capitals. C+: Congruent; C−: Incongruent; P+: With pitch accent; P−: Without pitch accent.

who did not participate in the functional magnetic resonance imaging (fMRI) study. These subjects were required to answer the question *Does the sentence make sense?* (Dutch: *Klopt de zin?*) by pressing a button as fast as possible: Green button (right index finger) for "yes", red button (right middle finger) for "no". We found that the detection rate for anomalies was significantly higher in the P+ condition (mean detection rate: 78.5%) than in the P− condition (mean detection rate: 63.6%; two-tailed paired *t*-test: $t_{(199)} = 7.93$, $P < 0.001$). For the 2 C+ conditions, no difference was found in false alarm rates (i.e. the rate of C+ sentences that were taken as anomalous) between the C+P+ (9%) and C+P− conditions (8%), as indicated by a two-tailed paired *t*-test: $t_{(199)} = 0.70$, $P = 0.49$.

We discarded 79 quartets of the 200 original quartets, on the grounds of the distribution of responses. To meet the first criterion of containing anomalies, we discarded quartets for which subjects more often judged the C−P+ version as correct than as incorrect (i.e. the anomaly was too subtle). We also discarded quartets for which subjects always noticed the anomaly in both C− versions (i.e., the anomaly was too obvious). To meet the second criterion of unproblematic C+ versions, we discarded quartets if the subjects judged the C+ versions more often as incorrect than as correct. We also discarded the quartets if the subjects judged the C− versions as incongruent more often than they judged the C+ versions as congruent. In the end, 121 quartets remained and we used 120 of them in the fMRI experiment (cf. Supplementary material). For these 120 items, the detection rate for anomalies was significantly higher in the P+ condition (mean detection rate: 81.5%) than that in the P− condition (mean detection rate: 57.7%; two-tailed paired *t*-test: $t_{(119)} = 9.42$, $P < 0.001$). For the 2 C+ conditions, no difference was found in false alarm rates between the C+P+ and C+P− conditions (7% and 6%, respectively; two-tailed paired *t*-test: $t_{(119)} = 0.77$, $P = 0.46$). We compared the frequency of the CWs in the congruent and incongruent conditions based on the Dutch SUBTLEX word corpus (Keuleers et al. 2010). Out of the 120 word pairs, 106 word pairs were registered in the corpus with log frequency (mean ± SD = 2.31 ± 0.95 and 2.25 ± 1.02 for the words in the congruent and incongruent conditions, respectively). A two-tailed *t*-test showed no frequency difference for word pairs ($t_{(105)} = 0.72$, $P = 0.48$).

Acoustic Analysis. To ensure that there was a measurable difference in prosodic pattern between CWs in P+ and P− conditions, we examined them with regards to acoustic measures that predict perceived prominence statistically. According to Streefkerk et al. (1999), perceived prominence in Dutch sentences that were read aloud could be predicted (with 81% accuracy) on the grounds of a combination of acoustic measurements, such as duration, intensity, and F0. In our analysis, we used similar measures to compare P+ experimental sentences to their P− equivalents. Furthermore, we compared the length of pauses after each CW, as pauses are known to highlight boundaries between words in Dutch (De Pijper and Sanderman 1994). We performed 6 *F*-tests, each with one of the following dependent variables: CW duration, CW intensity, mean F0 of each CW, standard deviation (std) for the F0 of each CW, root mean square (rms) for the amplitude of each CW, and length of any pause immediately following each CW. Table 2 presents the results for the acoustic measures and statistical analysis of the CWs. In P+ conditions, CWs had significantly longer durations, higher intensity, higher mean F0, larger F0 std, larger rms of amplitude, and longer pauses than in the P− conditions. Therefore, the acoustic measurements indicate a difference in perceived prominence of the 2 conditions. The duration of the sentences ranged from 3.5 to 7.1 s with an average duration of 5.4 s.

Experimental Lists. For the fMRI experiment, we made 4 lists by means of a Latin square procedure, with each list containing an equal number of items (30 items) per condition. No single participant listened to more than one version of a quartet, and all 4 versions were presented across 4 different experimental lists. In order to vary the position of pitch accent in sentences containing anomalies, we constructed 30 C− filler sentences. They contained an incongruent word, but the pitch accent was on one of the other words in the

Table 2
Acoustic measurements of critical words in target sentences

	P+ (with pitch accent)	P− (without pitch accent)	$F_{(1,119)}$
Duration (ms)	700 (196)	536 (190)	355***
Intensity (dB)	66 (2)	60 (3)	426***
F0 mean (Hz)	153 (14)	127 (13)	352***
F0 std	27 (8)	14 (12)	119***
Amplitude rms	0.04 (.01)	0.02 (.01)	363***
Pause (ms)	58 (7)	9 (2)	46***

Note: ***signifies significance at the 0.001 level. Means and standard deviations (in brackets) are presented.
std, standard deviation; rms, root mean square.

sentence. For example, in *Root crops like radish, CARROTS,* and *peas do not stand frost*, the word *peas* (which is not a root crop) was not realized with a pitch accent, whereas one of the other words, *carrots*, was. To have the same number of congruent and incongruent sentences on each list, we added 30 C+ fillers. These fillers were similar in format and content to C+P+ items. Consequently, there were 180 sentences in each experimental list (120 experimental items and 60 filler items). Each list was presented to the same number of subjects (6 subjects). For half of the list, we reversed the order of presentation. To assure that participants attended to the content of the sentences, 1 out of 6 trials contained a visually displayed comprehension question. The 30 comprehension questions were equally distributed over the 4 experimental and the 2 filler conditions. The comprehension question concerned the previous sentence and never referred to the CWs. Questions thus never referred to the anomalies. All comprehension questions were yes–no questions (half of which required a "yes" answer).

Procedure. Each trial started with a fixation cross lasting 3–7 s. Then a 300 ms auditory beep cue was presented to indicate the start of the upcoming sentence, and 700 ms later the subject heard a sentence. For 5 of 6 of the trials, the sentence presentation was immediately followed by the fixation cross, indicating a new trial. In the other one-sixths of the trials, a "yes/no" question was presented visually after a silent fixation period of 3–7 s. The question was presented for 3 s, during which subjects should indicate their answers by pressing a green button ("yes") using right index finger or a red button ("no") using right middle finger. Then the next trial began with the 3–7 s fixation. To minimize eye movements, we instructed subjects to look at a white fixation cross shown on the black screen throughout the experiment (except during questions). Before entering the scanner, the subjects completed a practice session containing 15 trials to familiarize them with the procedure and to ensure that they fully understood it.

fMRI Data Acquisition
Participants were scanned with a Siemens 3T Tim-Trio MR scanner, using a 32 channel surface coil. We acquired T_2*-weighted EPI-BOLD fMRI data using an ascending slice acquisition sequence (volume TR = 1.78 s, TE = 30 ms, 90° flip-angle, 31 slices, slice-matrix size = 64 × 64, slice thickness = 3 mm, slice gap = 0.5 mm, field of view (FOV) = 224 mm, voxel size = 3.5 × 3.5 × 3.0 mm). After the auditory spatial attention task (~20 min), subjects performed the spoken language task (~40 min). Finally, we acquired high-resolution anatomical MR images with a T_1-weighted 3D MPRAGE sequence (TR = 2300 ms, TE = 3.03 ms, 192 sagittal slices, slice thickness = 1.0 mm, voxel size = 1 × 1 × 1 mm, FOV = 256 mm).

Data Analysis

Preprocessing
The fMRI data were preprocessed using statistical parametric mapping 5 (SPM5) (Friston 2007). The first 5 images were discarded to avoid transient non-saturation effects. The functional images were

realigned, slice-time corrected, and then co-registered with the corresponding structural MR images using mutual information optimization. Subsequently, the images were spatially normalized (i.e. the normalized transformations were generated from the structural MR images and applied to the functional MR images), and transformed into a common anatomical space defined by the SPM Montreal Neurological Institute (MNI) T1 template. Finally, the normalized images were spatially smoothed using a 3D isotropic Gaussian kernel (FWHM = 8 mm).

Whole-Brain Analysis

The linear model used in the first-level single-subject analysis was based on the functional images acquired in the auditory spatial attention and spoken language task. The beta-images of the corresponding first-level regressors were used in the second-level random effects group analysis, separately for these 2 tasks.

In the first-level auditory spatial attention task analysis, the linear model included explanatory variables derived from the onsets and durations of different events: cue, filler tones, target, and end-of-trial "O". The target regressors were segregated on trial type: valid/invalid, left/right, and an absence of target. In addition, the trials with a response error were modeled separately. Regressors for the 6 realignment parameters were included for movement artifact correction as well as a high-pass filter (cutoff 128 s) to account for various low-frequency effects. The regressors (except the realignment and high-pass filter regressors) of the model were convolved with the canonical hemodynamic response function provided by SPM5. Then the beta-images of the "cue", "filler tones", "valid target" (the combination of valid cue trial/left side and valid cue trial/right side), "invalid target" (the combination of invalid cue trial/left side and invalid cue trial/right side), "target" (the combination of valid and invalid targets), and "no target" were generated. For the second-level random-effects analysis, the beta images were subjected to either 1-sample T-tests (for the cue and filler tones) or 1-way repeated-measures ANOVAs (for the contrasts invalid vs. valid target and target vs. no target).

The linear model for the first-level spoken language single-subject analysis included a regressor for the fixation, a regressor from the onset of the auditory beep cue to either the onset of the CW in the experimental conditions or to the offset of the filler sentences, condition regressors from the onset of the CW to the offset of the experimental sentences in the C+P+, C−P+, C+P−, C−P− conditions separately, and finally a regressor for the comprehension questions. Regressors for the 6 realignment parameters, as well as a high-pass filter (cutoff 128 s), were included in the model. The regressors (except the realignment and high-pass filter regressors) of the model were convolved with the canonical hemodynamic response function provided by SPM5. For the second-level statistical analysis, the beta-image related to the C+P+, C−P+, C+P−, C−P− conditions were used in a 2-way repeated-measures ANOVA: Congruence (Congruent: C+, Incongruent: C−), Prosodic pattern (With pitch accent: P+, Without pitch accent: P−). In addition, based on the strong a priori hypothesis that pitch accent will modulate the unification process, we also performed 2 1ne-way repeated-measures ANOVAs for the contrasts of C−P+ versus C+P− and C−P− versus C+P−.

For both tasks, the second-level statistical inference was based on the cluster-size statistics from the relevant second SPM[T] volumes and *P*-values corrected for multiple dependent comparisons (Friston 2007). SPMs were thresholded at the voxel level at $P < 0.005$ (uncorrected) to define clusters, and only clusters significant at $P < 0.05$ (family-wise error [FWE] corrected) are reported (unless otherwise specified). Local maxima within significant clusters are reported with their respective Z-values.

Results

Behavioral Results

For the auditory spatial attention task, all 24 included participants had a high correct response rate (>80%). Three other subjects made a high number of errors (correct response rate <65%) and were therefore excluded from further analysis. The

Table 3
Behavioral data showing validity effect in the spatial attention task

	Valid	Invalid
RT (ms)	414 (114)	449 (114)
Accuracy	95% (6%)	93% (10%)

Note: Means and standard deviations (in brackets) are presented.
RT, response time.

response time (RT) and accuracy for the 24 participants included in the auditory spatial attention task are summarized in Table 3.

A 1-way repeated-measures ANOVA was conducted for the RT and accuracy data separately, with validity of the cue (valid, invalid) as a within-subject factor. We found significantly shorter RT for the valid compared with the invalid condition ($F_{(1,23)} = 15.3$, $P < 0.001$), but no significant difference was found in the accuracy between the valid and invalid conditions ($F_{(1,23)} = 1.39$, $P = 0.25$).

In the spoken language task, the questions were designed to make sure that the participants attended to all sentences while lying in the scanner. We found that the participants had high response accuracy (mean accuracy 81%; standard deviation 12%) for 22 of the 30 questions. For these 22 questions, the incorrect responses and missed button presses were evenly distributed among the participants. For the remaining 8 questions, <20% of the participants were able to give correct answers. This was due to the difficulty of the questions rather than the absence of participants' attention, since it is unlikely that >80% of the participants were inattentive during the same item after the order of items had been randomized across participants. Therefore, the results showed that the participants attended to most of the sentences.

Brain Areas Activated in the Auditory Spatial Attention Task

During the cue period, we found a right superior and inferior parietal cluster ($P_{\text{FWE}} = 0.004$), which extended medially into the precuneus, paracentral, and postcentral cortex (shown in Fig. 2A). During the filler tones period, we found bilateral superior temporal clusters (left: $P_{\text{FWE}} < 0.001$; right: $P_{\text{FWE}} < 0.001$), extending into the inferior frontal and insular cortex on the left. In addition, posterior frontal cortex was activated (left: $P_{\text{FWE}} < 0.001$; right: $P_{\text{FWE}} < 0.001$), including bilateral supplementary motor areas, left precentral and postcentral cortex, as well as bilateral mid-anterior cingulate cortex (Fig. 2B). In the target period, the contrasts between the invalid and valid targets did not yield any significant difference. Therefore, we collapsed the 2 target conditions into 1 (including both valid and invalid targets) and compared it with the absence of target. This contrast yielded activations in bilateral perisylvian regions (left: $P_{\text{FWE}} < 0.001$; right: $P_{\text{FWE}} < 0.001$), including left precentral, postcentral and superior/inferior parietal cortex, anterior cingulate, insula, right inferior parietal cortex, and subcortical structures including the thalamus, as well as bilateral precuneus ($P_{\text{FWE}} = 0.01$) (see Fig. 2C). Table 4 displays the coordinates of local maxima for all the significant clusters.
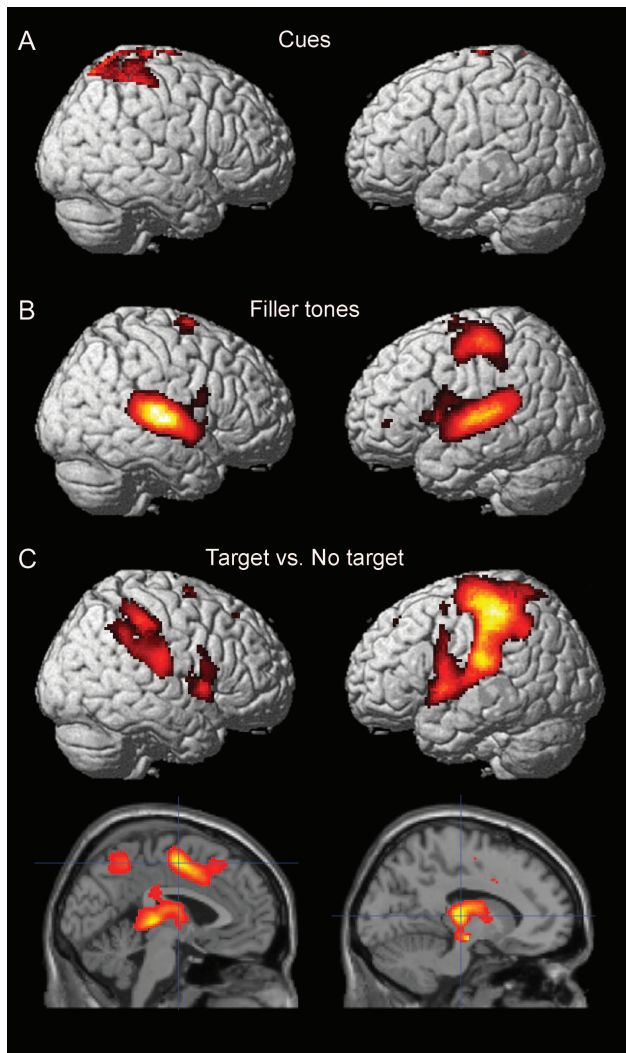
Figure 2. Significant brain activations found in the auditory spatial attention task (MNI stereotactic space; cluster-level $P_{\text{FWE-corrected}} < 0.05$; thresholded at the voxel-level $P_{\text{uncorrected}} < 0.005$). (A) Effect of attention orientation elicited by cues. The activated areas include right superior and inferior parietal cortex, extending medially into the precuneus, paracentral and postcentral cortex. (B) Effect of attention maintenance during the presentation of filler tones. The activated regions are bilateral superior temporal, extending to inferior frontal and insula, as well as posterior frontal cortex, including bilateral supplementary motor area, middle anterior cingulate, and left precentral and postcentral cortex. (C) Effect of bottom-up input, obtained by the contrast between target (including both valid and invalid targets) versus no target conditions. The presence of target activated bilateral perisylvian regions, including left precentral, postcentral, and superior/inferior parietal cortex, anterior cingulate, right inferior parietal cortex, bilateral precuneus, as well as subcortical area such as insula and thalamus. The sagittal slices are shown at $x = -2$ (bottom, left) with the cross-hair at the maximum in middle cingulate cortex and $x = -12$ (bottom, right) with the cross-hair at left thalamus.

Brain Areas Activated in the Spoken Language Task

In the spoken language task, the incongruent conditions elicited larger activations than the congruent conditions in the bilateral inferior and MFG (left: $P_{\text{FWE}} < 0.001$; right: $P_{\text{FWE}} = 0.005$), as well as left medial frontal region ($P_{\text{FWE}} = 0.03$; see Fig. 3A). Moreover, the separate analysis of the congruency effect for the pitch accent and no pitch accent conditions revealed that the incongruent sentences elicited larger LIFG activations in the pitch accent condition only ($P_{\text{FWE}} = 0.01$; see Fig. 3B). The comparison between conditions with and

without pitch accent on the CW revealed larger activations for the pitch accent condition in bilateral superior and middle temporal cortex (left: $P_{\text{FWE}} = 0.002$; right: $P_{\text{FWE}} < 0.001$), bilateral inferior parietal cortex (left: $P_{\text{FWE}} = 0.011$; right: $P_{\text{FWE}} = 0.002$), left inferior and middle frontal cortex ($P_{\text{FWE}} = 0.086$), as well as right inferior, middle, and posterior frontal cortex ($P_{\text{FWE}} = 0.009$; shown in Fig. 3C). The interaction between congruence and prosodic pattern showed a significant cluster ($P_{\text{FWE}} = 0.018$) in the right superior/inferior parietal and supramarginal cortex (see Fig. 3D). In order to test whether the interaction between congruence and prosodic pattern also exists in the left parietal cortex, we applied a small volume correction to the homologous region of the right parietal cortex. We defined a sphere of 10 mm radius centered on coordinates of the homologous activation peak in the right hemisphere (-34, -52, 56), within which the correction of multiple comparisons was performed. In this way, the sensitivity of the statistical analysis was improved. The results showed a significant cluster in the superior/inferior parietal cortex ($P_{\text{FWE}} = 0.033$, small volume correction). The peaks of activations are shown in Table 5.

To further describe the different activation patterns in the 4 conditions, we took the functional activation revealed in the interaction (right superior/inferior parietal and supramarginal cortex, and left superior/inferior parietal cortex) as regions of interest (ROIs). The average time course in the 4 conditions were separately calculated using MarsBaR (Brett et al. 2002). Since hemispherical differences are not within the focus of our analysis, and the 2 ROIs differ in size, we performed a 2-way repeated-measures ANOVA to the averaged beta values in each of the ROIs, without including Hemisphere as a factor. In both ROIs, we found a larger activation for the P+ than the P− condition ($F_{(1,23)} = 4.04$, $P = 0.056$; $F_{(1,23)} = 5.49$, $P = 0.028$, respectively for the right and left ROI), while no significant difference was found between the C− and the C+ condition ($F_{(1,23)} = 0.29$, $P = 0.60$; $F_{(1,23)} = 1.67$, $P = 0.21$, respectively, for the right and left ROI). Moreover, the difference between the P+ and P− conditions in the ROIs depended on the sentence congruence, as indicated by a significant interaction between the prosodic pattern and congruence ($F_{(1,23)} = 10.34$, $P = 0.004$; $F_{(1,23)} = 9.90$, $P = 0.005$, respectively, for the right and left ROI). A simple effect test showed that the difference between the P+ and P− conditions was only significant in the C− condition ($F_{(1,23)} = 20.8$, $P < 0.001$; $F_{(1,23)} = 18.2$, $P < 0.001$, respectively, for the right and left ROI), but not in the C+ condition ($F_{(1,23)} = 0.34$, $P = 0.56$; $F_{(1,23)} = 0.01$, $P = 0.93$, respectively, for the right and left ROI). Figure 4 displays the different activations among the 4 conditions in the ROIs. Since we are interested in comparing the relative contribution of each condition, we took the activation of C−P− condition, which exhibits the largest deactivation, as an arbitrary zero. The beta values of other conditions are relative values against that of the C−P− condition.

Common Areas Activated in Both the Auditory Spatial Attention Task and the Spoken Language Task

In order to determine whether the spoken language task and the auditory spatial attention task activated the same attention network, we overlaid the functional activations obtained in the auditory spatial attention task and the activations found in the contrast between the P+ and P− in the spoken language

Table 4

Results for the auditory spatial attention contrasts

Anatomical cluster	BA	Local maxima (x, y, z)	Cluster size	Cluster P_{FWE}	Voxel-level Z score
Cue period					
R superior parietal cluster			**779**	**0.004**	
Precuneus	7	(8, −50, 76)			3.01
Superior parietal cortex	7	(30, −64, 64)			3.39
Superior parietal cortex	7/40	(46, −46, 62)			3.55
Inferior parietal cortex	40	(46, −56, 58)			2.61
Paracentral cortex	6	(4, −18, 78)			3.88
Paracentral cortex	4	(6, −36, 78)			4.23
Postcentral	2	(52, −28, 60)			3.49
Postcentral	5/7	(38, −40,68)			3.63
Filler tones period					
L/R SMA-cingulate cluster			**2289**	**<0.001**	
L supplementary motor	6	(−8, −6, 64)			5.76
R supplementary motor	6	(4, 0, 62)			6.11
L mid-anterior cingulate cortex	24/32	(−10, 8, 34)			3.36
R mid-anterior cingulate cortex	24/32	(12, 8, 40)			3.27
L pre/postcentral cluster			**2165**	**<0.001**	
Precentral	4	(−36, −22, 60)			5.77
Precentral	6	(−32, −16, 70)			3.68
Postcentral	2	(−42, −24, 56)			5.54
L superior temporal cluster			**4737**	**<0.001**	
Heschl's gyri	41/42	(−48, −22, 6)			7.18
Superior temporal gyrus	22	(−52, −10, 2)			6.34
Superior temporal gyrus	22	(−62, −30, 10)			6.02
IFG	6/44	(−52, 6, 6)			3.27
Frontal operculum	44	(−48, 8, 4)			3.10
Superior-anterior insula	13/15/49	(−30, 20, 6)			4.53
Precentral	6	(−60, 2, 18)			3.57
R superior temporal cluster			**3846**	**<0.001**	
Heschl's gyri	41/42	(48, −22, 10)			7.13
superior temporal gyrus	22	(58, −18, 4)			7.83
Target versus no target					
L perisylvian and bilateral sub-cortical cluster			**17 318**	**<0.001**	
L precentral	4	(−36, −28, 70)			5.82
L postcentral	1/2/3	(−56, −24, 40)			5.84
L postcentral	2	(−42, −32, 58)			6.56
L superior parietal cortex	2/40	(−56, −32, 54)			6.61
L supramarginal cortex	40	(−54, −22, 22)			6.44
L superior temporal	42	(−50, −14, 8)			3.84
L insula	14/16	(−46, 0, 8)			5.88
L middle frontal cortex	6/44	(−60, 6, 24)			4.20
L putman	—	(−30, −2, −4)			4.12
superior colliculus	—	(0, −28, 4)			5.51
R caudate	—	(14, 0, 16)			3.99
R thalamus	—	(14, −8, 12)			4.83
R anterior cingulate	32	(4, 16, 40)			3.88
R insula	14/16	(44, 6, 2)			4.11
L/R precuneus cluster			**602**	**0.01**	
L precuneus	7	(−2, −52, 54)			3.74
R precuneus	7	(4, −62, 56)			3.69
R supramarginal cluster			**1968**	**<0.001**	
R inferior parietal cortex	40	(56, −36, 52)			4.05
R inferior parietal cortex	40	(54, −56, 44)			2.72
R inferior parietal cortex	40	(44, −46, 38)			2.61
R supramarginal cortex	40	(54, −42, 38)			3.30
R supramarginal cortex	40	(66, −28, 30)			3.71
R supramarginal cortex	40	(64, −18, 20)			4.81

Note: The significant clusters were obtained under a threshold of $P_{uncorrected} < 0.005$ at the voxel level, and $P_{FWE-corrected} < 0.05$ at the cluster level. BA = Brodmann's area; x, y, z = the original SPM x, y, z coordinates in millimeters of the MNI space; anatomical labels are derived from the Automated Anatomical Labeling map (AAL, Tzourio-Mazoyer et al. 2002) and from Brodmann's atlas in MRICron. The rows in boldface indicate a maximum in the significant clusters.

task. We chose this approach for 2 reasons: First, it is a conservative approach, as the activations of both tasks reached significance in the whole-brain analysis. Secondly, as the models constructed in the 2 tasks are very different, it would not have been feasible to carry out a direct conjunction analysis in the same model. Figure 5*A* shows the activations obtained from both tasks as well as the overlap between them. We found that these 2 tasks activated some common regions, including bilateral superior temporal gyrus, bilateral inferior parietal cortex (extending into the IPS), as well as left precentral cortex. In addition, Figure 5*B* shows the regions that were activated in both the auditory spatial attention task and the regions showing interaction effects in the spoken language task.

Discussion

This study aimed to examine the neural correlates of (prosodic) IS marking during sentence comprehension. In particular, we were interested in whether a common attention network was recruited for both linguistic, IS induced, attention and for non-linguistic attention in relation to auditory spatial processing. In a spoken language task, we indepen-
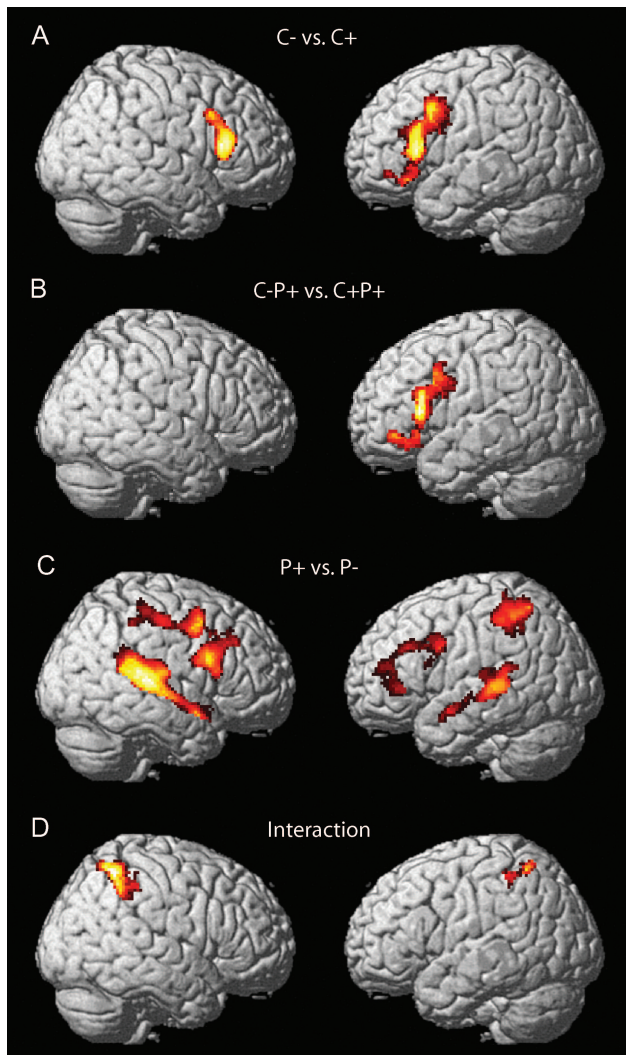
Figure 3. Significant brain activations in the spoken language task (MNI stereotactic space; cluster-level $P_{FWE-corrected} < 0.05$, except for the left frontal cluster in contrast B: $P_{FWE-corrected} = 0.083$; thresholded at the voxel-level $P_{uncorrected} < 0.005$). (*A*) Effect of congruence. The incongruent sentences had stronger activations in bilateral inferior and MFG, as well as left medial frontal region. (*B*) Effect of congruency in the pitch accent condition. The incongruent sentences elicited stronger activations in left IFG. No such activation was found in the no pitch accent condition. (*C*) Effect of the prosodic pattern. Relative to the no pitch accent condition, the pitch accent condition strongly activated bilateral superior and middle temporal cortex, bilateral inferior parietal cortex, left inferior and middle frontal cortex, as well as right inferior, middle, and posterior frontal cortex. (*D*) The interaction between congruence and prosodic pattern. The activations included bilateral superior/inferior parietal cortex and right supramarginal cortex. Note that the activation in the left parietal lobe was obtained after small volume correction.

dently manipulated the prosodic pattern and congruence of sentences. The prosodic manipulation showed bilateral involvement of the inferior parietal, superior, and middle temporal, as well as inferior, middle, and posterior parts of the frontal cortex. We compared the activations caused by pitch accent in the spoken language task with activations from an auditory spatial attention task. Both tasks activated a common attention network involving bilateral inferior parietal, superior temporal, and left precentral cortex. In addition, we observed an interaction between the prosodic pattern and congruence in bilateral inferior parietal regions: for

incongruent sentences, but not for congruent ones, there was a larger activation if the incongruent word was realized with a pitch accent, than without pitch accent. Finally, incongruent sentences affected bilateral inferior and MFG. We discuss the results in more detail below.

Neural Correlates of IS Marking Overlap with Attention Networks

In the spoken language task, a comparison of conditions with prosodic alternations revealed larger activations for the pitch accent condition in bilateral inferior parietal cortex, superior and middle temporal cortex, inferior and middle frontal cortex as well as right posterior frontal cortex. The auditory spatial attention task activated both dorsal and ventral attention networks (Corbetta et al. 2000, 2002). We compared the activations in these 2 tasks and found that some regions were overlapping. The shared network included the bilateral superior temporal gyrus, inferior parietal cortex (extending into the IPS), and the left precentral cortex. The existence of a shared network indicates that linguistic attention is not separate from non-linguistic attention.

As mentioned in the Introduction, the inferior parietal cortex including IPS and dorsal frontal cortex play a role in deploying attention to task-relevant features or locations of objects (Corbetta et al. 2008). In general, the IPS is not involved in language processing (Vigneau et al. 2006). Only when a language-processing task focuses the attention of the listener, we might see the involvement of this region. For instance, previous studies have found inferior parietal activation for tasks where subjects were explicitly asked to discriminate between prosodic structures: to identify the placement of a contrastive stress (Tong et al. 2005; Perrone et al. 2010) and to discriminate sentence intonation (Tong et al. 2005). Unlike these studies, our study did not explicitly ask subjects to attend to prosody but to comprehend sentences. Therefore, we take the activations in the attention-related regions in our language task as induced by prosody itself, not by external factors such as specific instructions. Note that the resulting attention effects are inseparable from acoustic energy effects in this study, as the IS differences in our design involved manipulation of prosodic prominence. Still, the activation of the superior/inferior parietal cortex is not likely to be purely driven by the processing of increased acoustical energy, as an interaction between Congruence and Prosodic pattern was revealed in these regions. Besides, literature on attention in the visual spatial domain have also identified networks in the parietal and frontal cortex (Corbetta et al. 2000). Nevertheless, the overlap in bilateral superior temporal gyrus can both be interpreted as related to general attention (Corbetta et al. 2000; Mayer et al. 2006; Hill and Miller 2010) and as related to acoustic analysis of prominent sounds (Emmorey et al. 2003).

Overall, the overlap in the employed attention networks suggests that IS-marking (marked by prosody in the present study) modulated a domain general attention network. Pitch accent signaled the saliency of the focused words and thereby recruited attentional resources for increased processing. The involvement of both dorsal and ventral attention networks indicate that IS allocates attentional resources in both top-down and bottom-up manners. Although a general attention network has been proposed by a number of studies

Table 5
Results for the language contrasts

Anatomical cluster	BA	Local maxima (x, y, z)	Cluster size	Cluster P_{FWE}	Voxel-level Z score
Main effect of congruence: C− vs. C+					
L inferior/middle frontal cluster			**1943**	**<0.001**	
IFG	47	(−44, 42,−6)			2.92
IFG	45	(−40, 20, 20)			3.61
IFG	44	(−40, 8, 30)			3.71
anterior insula	13/15/49	(−32, 26, −4)			3.68
MFG	9	(−44, 10, 38)			3.53
MFG	46	(−52, 32, 22)			3.16
L medial frontal cluster	**6/8**	**(−6, 28, 54)**	**580**	**0.030**	**4.09**
R inferior/middle frontal cluster			**860**	**0.005**	
IFG	45	(50, 24, 12)			4.13
middle/IFG	45/46	(46, 22, 28)			3.43
middle/IFG	9/44	(42, 16, 32)			3.75
Congruency effect of the P+ condition: C− P+ versus C+P+					
L inferior frontal cluster			**1239**	**0.001**	
left IFG	44	(−40, 6, 30)			3.69
Left IFG	45	(−44, 18, 10)			3.51
left IFG	47	(−40, 38, −6)			3.36
Main effect of prosodic pattern: P+ versus P−					
L frontal cluster			**437**	**0.086**	
IFG	44	(−60, 14, 22)			3.03
IFG	45/46	(−52, 40, 6)			3.33
MFG	46	(−44, 46, 18)			3.42
Superior frontal/precentral gyrus	6/9/44	(−56, 10, 36)			3.60
L temporal cluster			**998**	**0.002**	
Middle temporal gyrus	21	(−52, −30, 0)			3.84
Middle temporal gyrus	21/22	(−40, −2, −14)			3.57
Superior temporal gyrus	22	(−60, −42, 16)			2.90
Superior temporal gyrus	22/42	(−62, −24, 12)			2.88
L parietal cluster			**729**	**0.011**	
Postcentral gyrus/superior parietal	2/7	(−36, −38, 68)			2.85
Superior parietal cortex	7	(−40, −50, 60)			3.81
Inferior parietal cortex	40	(−54, −40, 52)			3.57
Supramarginal gyrus	40	(−38, −50, 40)			2.95
R inferior/middle frontal cluster			**761**	**0.009**	
IFG	44/45	(50, 12, 24)			3.98
MFG	9	(48, 32, 34)			3.17
R middle/posterior frontal cluster			**1018**	**0.002**	
MFG	6/8/9	(30, 8, 44)			3.72
Precentral gyrus	4/6	(46, 4, 44)			3.50
Postcentral gyrus	1/2/3	(54, −20, 50)			3.54
Supramarginal gyrus	40	(48, −36, 46)			2.98
R temporal cluster			**2468**	**<0.001**	
Superior/middle temporal gyrus	21/22	(58, 6, −14)			4.25
Superior temporal gyrus	22	(60, −30, 8)			4.86
Superior temporal gyrus	41/42	(46, −36, 10)			4.69
Supramarginal gyrus	40	(52, −16, 14)			3.07
Interaction between congruence and prosodic pattern					
R parietal cluster			**656**	**0.018**	
Superior parietal cortex	7	(24, −58, 60)			3.04
Inferior parietal cortex	7/40	(42, −48, 48)			3.40
Supramarginal gyrus	40	(40, −38, 44)			3.26
L parietal cluster			**72**	**0.033**[SVC]	
Superior parietal cortex	7	(−30, −54, 58)			3.13
Inferior Parietal cortex	7/40	(−40, −46, 56)			2.82

Note: The significant clusters were obtained under a threshold of $P_{uncorrected} < 0.005$ at the voxel level, and $P_{FWEcorrected} < 0.05$ at the cluster level. The rows in boldface indicate a maximum in the significant clusters.

BA, Brodmann's area; x, y, z = the original SPM x, y, z coordinates in millimeters of the MNI space; anatomical labels are derived from the Automated Anatomical Labeling map (AAL, Tzourio-Mazoyer et al. 2002) and from Brodmann's atlas in MRICron. SCV, small volume correction.

(Wildgruber et al. 2004; Shomstein and Yantis 2006; Osaka et al. 2007; Hill and Miller 2010), few direct comparisons have been conducted between linguistic and non-linguistic tasks. Our results demonstrate that there is a common attention network for voluntarily orienting attention to a particular feature or location, whether linguistic or non-linguistic.

In addition to the overlapping brain regions in the 2 tasks, the prosodic alternation also modulated the brain regions in the bilateral inferior and middle frontal cortex. Although these regions have been associated with attention (for a review, see Corbetta et al. 2008), they were not activated in our auditory spatial attention task. The engaged brain regions can be seen as associated with unification of the linguistic input (Hagoort 2005; Hagoort et al. 2009). Activations related to specialized linguistic functions were also reported by Cristescu et al. (2006). In a cued lexical-decision task, they compared brain activations of spatial orienting (with a cue to the position of a target word) with that of semantic orienting (with a cue to the semantic category of the target word). They found that besides a common activation of fronto-parietal
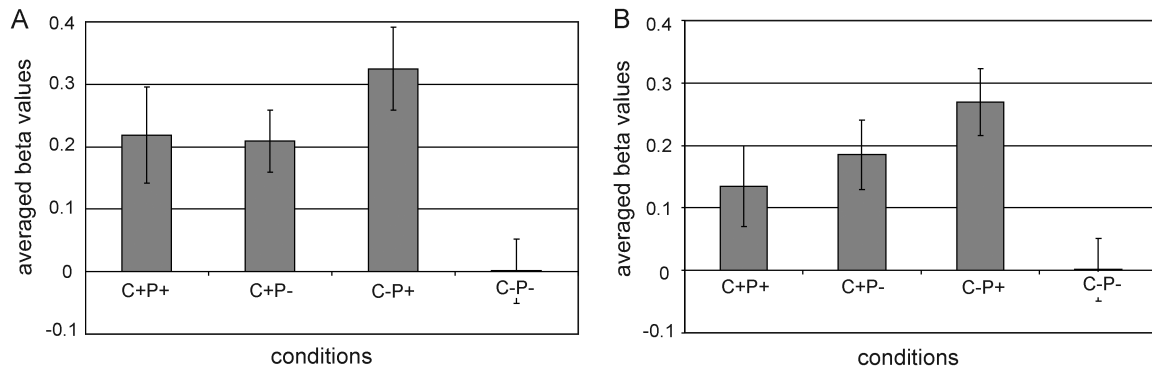
Figure 4. Different activations in the 4 conditions in the ROI of (*A*) the left superior/inferior parietal cortex and (*B*) the right superior/inferior parietal and right supramarginal region. The gray bars represent the averaged beta values of 4 conditions in the ROIs after scaling based on the activation in the C−P− condition: The activation in the C−P− condition was taken as an arbitrary zero in the diagram, and the magnitudes in the others conditions were relative values compared with that of C−P−. The vertical lines indicate the standard error for each condition. C+P+: Congruent, with pitch accent; C+P−: Congruent, without pitch accent; C−P+: Incongruent, with pitch accent; C−P−: Incongruent, without pitch accent.

networks, orienting to a semantic category selectively activated brain areas associated with semantic analysis of words, such as the left anterior inferior frontal cortex. This suggests that the type of attended information partly determines the activation of brain regions beyond a domain general attention network.

An Interaction in Inferior Parietal Cortex: The Activation of the Attention Network Depends on Sentence Congruence

In the spoken language task, we found an interaction between congruence and prosodic pattern in the right superior/inferior parietal (including IPS) and supramarginal cortex, as well as the left inferior parietal cortex. These regions were part of the attention network. For congruent sentences, the activation of these regions was not affected by the prosodic alternation. However, for the incongruent sentences, there was more activation in these regions in the case of an incongruent word with pitch accent compared with an incongruent word without pitch accent. Bisley and Goldberg (2010) proposed that IPS plays a role in prioritizing where to direct attention. For our spoken language task, both congruence (affecting unification) and accentuation (affecting perception) can be seen as important to prioritizing where attention is directed. In the congruent conditions, prosody has little influence on the activation of the attention networks, as the congruent words can be integrated in the sentential context without difficulties. This unification process requires the same amount of attention irrespective of the prosodic pattern of the sentences, resulting in similar levels of parietal activation. In contrast, in the incongruent conditions, the prosodic focus strongly activated the attention network due to the increased processing complexity. However, the anomalous words without pitch accent were not prosodically marked and showed the lowest parietal activation. They can be seen as having low priority compared with CWs in other conditions, as they were neither prosodically marked, nor congruent. These incongruent words without pitch accent can be taken as irrelevant errors which were disregarded and received little attention. Contrary to the incongruent words without pitch accent (in the C−P− condition), the congruent words without pitch accent (in the C+P− condition) were taken as relevant to unification, although they were not

prosodically prominent. This led to larger activations for the words in the C+P+ condition. In short, most attention was allocated to the anomalous words with pitch accent, and the inferior parietal cortex was activated most strongly for these words. Overall, the interaction effect reflects a bidirectional information exchange between the processing of linguistic content and the attention network.

It might be argued that the observed interaction in the parietal cortex simply reflects the greater effort or greater success in identifying the anomaly when it carried a pitch accent, but this argument is not necessarily inconsistent with the attention account, as the effort or success could be the consequence of attention deployment in the IPL.

Unification of Incongruent Sentences Activated the Frontal Cortex

The incongruent sentences elicited larger activations than the congruent sentences in bilateral inferior and MFG, as well as medial frontal region. In the light of previous work on semantic and pragmatic anomalies, it is not surprising to find LIFG activations in response to anomalies, as it reflects the increased unification load for incongruent sentences (for a review, see Hagoort et al. 2009). The involvement of the right IFG has been related to the construction of a discourse model (Menenti et al. 2009; Tesink et al. 2009), while the medial frontal region might reflect re-evaluation of the plausibility of the sentences (Stowe et al. 2005; Zhu et al. 2009).

Considering the sensitivity of LIFG to unification and the IS modulation on the semantic/pragmatic unification, we expected that the pitch accent would modulate the activity in LIFG in response to anomalies. We predicted a greater LIFG activation for anomalies with pitch accent than for those without pitch accent. Although no interaction between congruence and prosodic pattern was revealed in the LIFG, separate analysis of the congruency effect for the pitch accent and no pitch accent conditions confirmed our prediction. The LIFG activation pattern is consistent with that in the parietal cortex. In the pitch accent conditions, more attention was allocated to the anomalous words, leading to an increase of processing (unification) load, resulting in the LIFG activation. However, a less amount of attention was directed toward the anomalous words without pitch accent. The difference in level of attention between the 2 conditions might
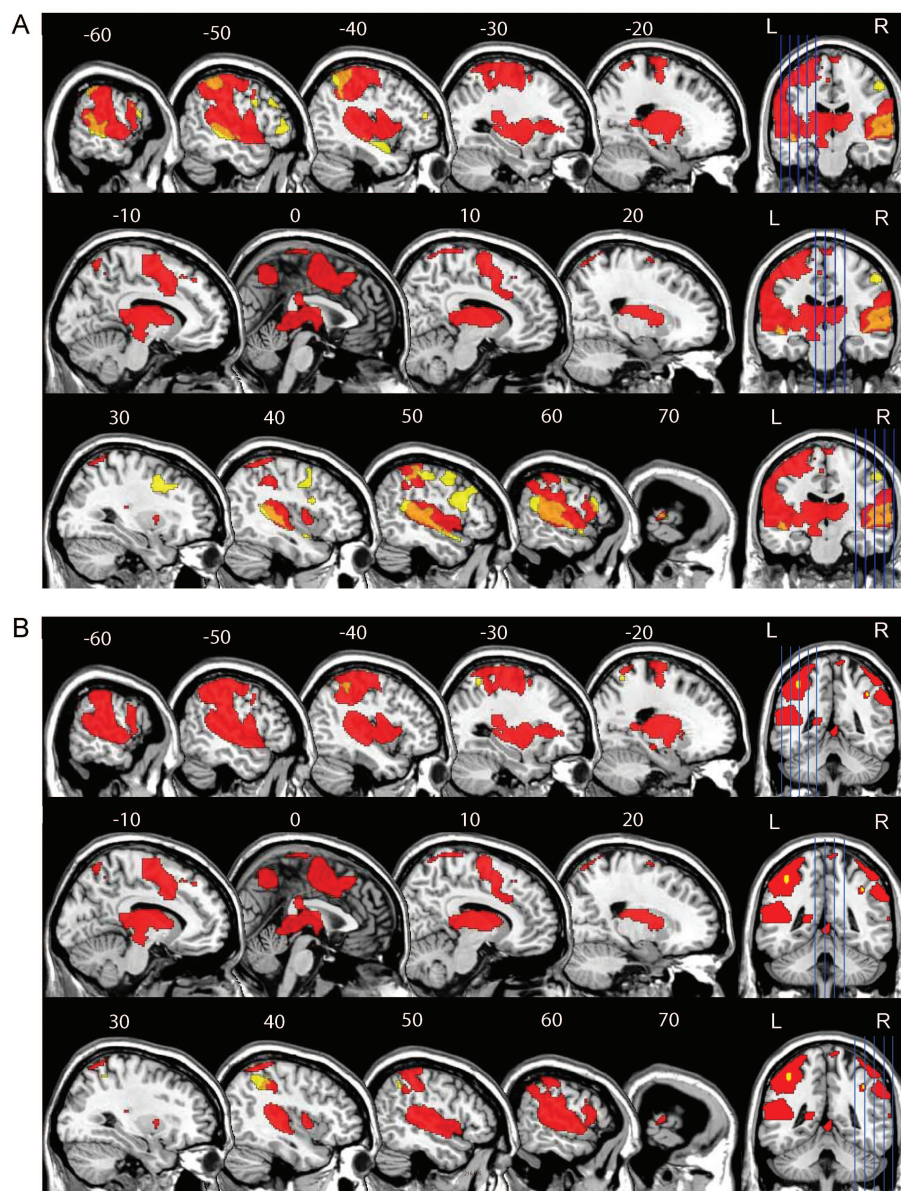
Figure 5. (A) The regions that were activated in both the auditory spatial attention task (in red) and the contrast between P+ (with pitch accent) and P− (without pitch accent) in the spoken language task (in yellow). The overlap (in orange) includes bilateral superior temporal gyrus, bilateral inferior parietal cortex (extending into the intraparietal sulcus), as well as left precentral cortex. (B) The regions that were activated in both the auditory spatial attention task (in red) and the regions showed an interaction between Prosodic pattern and Congruence in the spoken language task (in yellow). The overlap (in orange) includes bilateral superior/inferior parietal cortex. The activations were shown in multiple sagittal slices (x-coordinate, in equal 10 mm intervals, from −60 to 70 mm). The coronal slice is shown at y = 6 and −44, respectively, for (A) and (B). (MNI stereotactic space; cluster-level $P_{\text{FWE-corrected}} < 0.10$; thresholded at the voxel-level $P_{\text{uncorrected}} < 0.005$).

explain the difference in the level of activation of LIFG. Given that increased LIFG activation can be interpreted as increased processing/unification load, the anomalies for words without pitch accent were less attended to and shallowly processed and therefore did not result in increased activation of LIFG.

These results are in line with our behavioral pre-test which showed a higher frequency of semantic illusions when the incongruent critical word was realized without pitch accent (42%) than with pitch accent (19%). However, this difference in the anomaly detection rate between conditions with and without pitch accent may in fact have been too subtle to result in a significant interaction effect in LIFG.

The Auditory Spatial Attention Task Activated Both Dorsal and Ventral Attention Networks

During the cue period of the auditory spatial attention task, the dorsal parietal cortex was activated, including right superior and inferior parietal, precuneus, paracentral, and postcentral cortex. These regions are associated with top-down attention (Corbetta et al. 2000, 2002). During the filler tone period, we found activations in the bilateral superior temporal cortex, the left superior frontal cortex (precentral and postcentral cortex), the bilateral supplementary motor cortex, and the bilateral cingulate cortex. The superior frontal cortex constitutes a part of the top-down attention network. The activation of bilateral temporal cortex is likely to relate to the acoustic input from the

filler tones, which were used to increase the task difficulty. The activation of the bilateral supplementary motor cortex may relate to response preparation (Corbetta and Shulman 2002) and the anterior cingulate activation is associated with task-control (Dosenbach et al. 2006). Overall, the activations in the cue and filler tone periods largely activated the dorsal attention network.

The target activated both the dorsal top-down (left precentral, postcentral, IPS, precuneus) and the ventral bottom-up (bilateral TPJ and supramarginal cortex) network, as well as the thalamus. The ventral network activation relates to target detection, while the recruitment of the dorsal network indicates an interplay between the 2 attention networks (Shulman et al. 2003; Corbetta et al. 2008). Unlike Corbetta et al. (2000), we did not find significant differences between the invalid target and the valid target trials.

Conclusions

By comparing the brain activations elicited by prosodic IS marking with the attention networks localized in an auditory spatial attention task, we have demonstrated that IS, as a linguistic device, recruits a domain-general attention network. This network includes superior/inferior parietal cortex (extending to IPS and postcentral gyrus) and left precentral cortex. The activation of this attention network is sensitive to the semantic/pragmatic congruence of language input, with the strongest activation when the anomalous word is accented. The results suggest that language comprehension recruits a domain general attention network based on linguistic devices, such as pitch accent that marks the focus of an expression. Therefore, attention and language comprehension appear to be highly interactive at a neurobiological level.

Supplementary Material

Supplementary material can be found at: http://www.cercor.oxfordjournals.org/.

Funding

Notes

Conflict of Interest: None declared.

References

Beckman ME. 1996. The parsing of prosody. Lang Cognitive Proc. 11 (1):17–68.

Birch S, Rayner K. 1997. Linguistic focus affects eye movements during reading. Mem Cognition. 25(5):653–660.

Bisley JW, Goldberg ME. 2010. Attention, intention, and priority in the parietal lobe. Annu Rev Neurosci. 33(1):1–21.

Boersma P, Weenink D. 2002. Praat 4.0: A system for doing phonetics with the computer [computer software]. Amsterdam (the Netherlands): Universiteit van Amsterdam.

Bredart S, Docquier M. 1989. The Moses illusion: A follow-up on the focalization effect. Curr Psychol Cogn. 9:357–362.

Bredart S, Modolo K. 1988. Moses strikes again: Focalization effect on a semantic illusion. Acta Psychol. 67(2):135–144.

Brett M, Anton JL, Valabregue R, Poline JB. 2002. Region of interest analysis using an SPM toolbox. In: the 8th International Conference on Functional Mapping of the Human Brain, Sendai, Japan.

Büring D. 2007. Semantics, intonation and information structure. In: Gillian R, Charles R, editors. The Oxford handbook of linguistic interfaces. Oxford (United Kingdom): Oxford University Press. p. 445–476.

Corbetta M, Kincade JM, Ollinger JM, McAvoy MP, Shulman GL. 2000. Voluntary orienting is dissociated from target detection in human posterior parietal cortex. Nat Neurosci. 3(3):292–297.

Corbetta M, Kincade JM, Shulman GL. 2002. Neural systems for visual orienting and their relationships to spatial working memory. J Cognitive Neurosci. 14(3):508–523.

Corbetta M, Patel G, Shulman GL. 2008. The reorienting system of the human brain: From environment to theory of mind. Neuron. 58 (3):306–324.

Corbetta M, Shulman GL. 2002. Control of goal-directed and stimulus-driven attention in the brain. Nat Rev Neurosci. 3(3):201–215.

Cristescu TC, Devlin JT, Nobre AC. 2006. Orienting attention to semantic categories. Neuroimage. 33(4):1178–1187.

Cutler A, Fodor JA. 1979. Semantic focus and sentence comprehension. Cognition. 7(1):49–59.

De Pijper JR, Sanderman AA. 1994. On the perceptual strength of prosodic boundaries and its relation to suprasegmental cues. J Acoust Soc Am. 96(4):2037–2047.

Dosenbach NUF, Visscher KM, Palmer ED, Miezin FM, Wenger KK, Kang HC, Burgund ED, Grimes AL, Schlaggar BL, Petersen SE. 2006. A core system for the implementation of task sets. Neuron. 50(5):799–812.

Emmorey K, Allen JS, Bruss J, Schenker N, Damasio H. 2003. A morphometric analysis of auditory brain regions in congenitally deaf adults. Proc Natl Acad Sci United States of America. 100 (17):10049–10054.

Erickson TD, Mattson ME. 1981. From words to meaning: A semantic illusion. J Verbal Lear Verbal Behav. 20(5):540–551.

Ferreira F, Ferraro V, Bailey KGD. 2002. Good-enough representations in language comprehension. Curr Dir Psychol Sci. 11:11–15.

Frazier L, Clifton C. 1997. Construal: Overview, motivation, and some new evidence. J Psycholinguist Res. 26:277–295.

Friston K. 2007. Statistical parametric mapping. In: Friston K, Ashburner J, Kiebel S, Nichols T, Penny W, editors. Statistical parametric mapping: the analysis of functional brain images. Amsterdam (the Netherlands): Academic Press. p. 10–31.

Gussenhoven C. 2008. Notions and subnotions in information structure. Acta Linguistica Hung. 55(3–4):381–395.

Hagoort P. 2005. On Broca, brain, and binding: A new framework. Trends Cogn Sci. 9(9):416–423.

Hagoort P, Baggio G, Willems RM. 2009. Semantic unification. In: Gazzaniga MS, editor. The new cognitive neurosciences. Cambridge: MIT Press. p. 819–836.

Hagoort P, Hald L, Bastiaansen M, Petersson KM. 2004. Integration of word meaning and world knowledge in language comprehension. Science. 304(5669):438–441.

Halliday MAK. 1967. Notes on transitivity and theme in English, part II. J Linguist. 3:199–244.

Hill KT, Miller LM. 2010. Auditory attentional control and selection during cocktail party listening. Cereb Cortex. 20(3):583–590.

Hornby PA. 1974. Surface structure and presupposition. J Verbal Lear Verbal Behav. 13(5):530–538.

Keuleers E, Brysbaert M, New B. 2010. SUBTLEX-NL: A new measure for Dutch word frequency based on film subtitles. Behav Res Meth. 42(3):643–650.

Kotschi T. 2006. Information structure in spoken discourse. In: Keith Brown, editor. The encyclopedia of language and linguistics. 2nd ed. Oxford: Elsevier. p. 677–682.

Li X, Gandour J, Talavage T, Wong D, Dzemidzic M, Lowe M, Tong Y. 2003. Selective attention to lexical tones recruits left dorsal frontoparietal network. Neuroreport. 14(17):2263–2266.

Li X, Gandour JT, Talavage T, Wong D, Hoffa A, Lowe M, Dzemidzic M. 2010. Hemispheric asymmetries in phonological processing of tones versus segmental units. Neuroreport. 21 (10):690–694.

Mayer AR, Harrington D, Adair JC, Lee R. 2006. The neural networks underlying endogenous auditory covert orienting and reorienting. Neuroimage. 30(3):938–949.

Menenti L, Petersson KM, Scheeringa R, Hagoort P. 2009. When elephants fly: Differential sensitivity of right and left inferior frontal gyri to discourse and world knowledge. J Cognitive Neurosci. 21 (12):2358–2368.

Meyer M, Alter K, Friederici A. 2003. Functional MR imaging exposes differential brain responses to syntax and prosody during auditory sentence comprehension. J Neurolinguist. 16(4–5):277–300.

Miller N, Lowit A, O'Sullivan H. 2006. What makes acquired foreign accent syndrome foreign? J Neurolinguist. 19(5):385–409.

Osaka M, Komori M, Morishita M, Osaka N. 2007. Neural bases of focusing attention in working memory: An fMRI study based on group differences. Cogn Affect Behav Neurosci. 7:130–139.

Perrone M, Dohen M, Loevenbruck H, Sato M, Pichat C, Yvert G, Baciu M. 2010. An fMRI study of the perception of contrastive prosodic focus in French. Speech Prosody. 100506(1-4).

Salmi J, Rinne T, Degerman A, Salonen O, Alho K. 2007. Orienting and maintenance of spatial attention in audition and vision: multimodal and modality-specific brain activations. Brain Struct Funct. 212(2):181–194.

Shipp S. 2004. The brain circuitry of attention. Trends Cogn Sci. 8 (5):223–230.

Shomstein S, Yantis S. 2006. Parietal cortex mediates voluntary control of spatial and nonspatial auditory attention. J Neurosci. 26 (2):435–439.

Shulman GL, McAvoy MP, Cowan MC, Astafiev SV, Tansy AP, d'Avossa G, Corbetta M. 2003. Quantitative analysis of attention and detection signals during visual search. J Neurophysiol. 90(5):3384–3397.

Stowe LA, Haverkort M, Zwarts F. 2005. Rethinking the neurological basis of language. Lingua. 115(7):997–1042.

Streefkerk BM, Pols LCW, Bosch LFMT. 1999. Acoustical features as predictors for prominence in read aloud Dutch sentences used in ANN's. P Eurospeech'99. 1:551–554.

Tesink CMJY, Petersson KM, van Berkum JJA, van den Brink D, Buitelaar JK, Hagoort P. 2009. Unification of speaker and meaning in language comprehension: An fMRI study. J Cognitive Neurosci. 21 (11):2085–2099.

Tong Y, Gandour J, Talavage T, Wong D, Dzemidzic M, Xu Y, Li X, Lowe M. 2005. Neural circuitry underlying sentence-level linguistic prosody. NeuroImage. 28(2):417–428.

Trueswell JC, Tanenhaus MK, Garnsey SM. 1994. Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. J Mem Lang. 33(3):285–318.

Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, Mazoyer B, Joliot M. 2002. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage. 15 (1):273–289.

van Leeuwen T, Lamers M, Petersson KM, Gussenhoven C, Rietveld T, Poser B, Hagoort P. Submitted. Prosody and information structure: An fMRI study.

Vigneau M, Beaucousin V, Hervé PY, Duffau H, Crivello F, Houdé O, Mazoyer B, Tzourio-Mazoyer N. 2006. Meta-analyzing left hemisphere language areas: Phonology, semantics, and sentence processing. NeuroImage. 30(4):1414–1432.

Vossel S, Thiel CM, Fink GR. 2006. Cue validity modulates the neural correlates of covert endogenous orienting of attention in parietal and frontal cortex. NeuroImage. 32(3):1257–1264.

Wang L, Bastiaansen M, Yang Y, Hagoort P. 2011. The influence of information structure on the depth of semantic processing: How focus and pitch accent determine the size of the N400 effect. Neuropsychologia. 49(5):813–820.

Wang L, Bastiaansen M, Yang Y, Hagoort P. in revision. Information structure influences depth of syntactic processing: Event-related potential evidence for the Chomsky illusion.

Wang L, Hagoort P, Yang Y. 2009. Semantic illusion depends on information structure: ERP evidence. Brain Res. 1282:50–56.

Wildgruber D, Hertrich I, Riecker A, Erb M, Anders S, Grodd W, Ackermann H. 2004. Distinct frontal regions subserve evaluation of linguistic and emotional aspects of speech intonation. Cereb Cortex. 14(12):1384–1389.

Zhu Z, Zhang JX, Wang S, Xiao Z, Huang J, Chen HC. 2009. Involvement of left inferior frontal gyrus in sentence-level semantic integration. NeuroImage. 47(2):756–763.