

# Functional Integration across Brain Regions Improves Speech Perception under Adverse Listening Conditions

Jonas Obleser,<sup>1</sup> Richard J. S. Wise,<sup>2,3</sup> M. Alex Dresner,<sup>3</sup> and Sophie K. Scott<sup>1,3</sup>

<sup>1</sup>Institute of Cognitive Neuroscience, University College London, London WC1N 3AR, United Kingdom, and <sup>2</sup>Division of Neuroscience and Mental Health and <sup>3</sup>Medical Research Council Clinical Sciences Centre, Imperial College, London W12 0NN, United Kingdom

Speech perception is supported by both acoustic signal decomposition and semantic context. This study, using event-related functional magnetic resonance imaging, investigated the neural basis of this interaction with two speech manipulations, one acoustic (spectral degradation) and the other cognitive (semantic predictability). High compared with low predictability resulted in the greatest improvement in comprehension at an intermediate level of degradation, and this was associated with increased activity in the left angular gyrus, the medial and left lateral prefrontal cortices, and the posterior cingulate gyrus. Functional connectivity between these regions was also increased, particularly with respect to the left angular gyrus. In contrast, activity in both superior temporal sulci and the left inferior frontal gyrus correlated with the amount of spectral detail in the speech signal, regardless of predictability. These results demonstrate that increasing functional connectivity between high-order cortical areas, remote from the auditory cortex, facilitates speech comprehension when the clarity of speech is reduced.

**Key words:** speech; semantics; auditory cortex; prefrontal cortex; functional connectivity; fMRI

## Introduction

Everyday speech perception is successful despite listening conditions that are usually less than ideal (e.g., the speech signal may be degraded by being embedded in ambient noise or echoes or may be subject to spectral reductions or compression down telephone lines). Despite these distortions, the listener is usually unaware of any difficulty in understanding what has been said. It has been recognized since the 1950s that an important factor supporting speech comprehension is the semantic context in which it is heard. Thus, words embedded in sentences are usually better understood than isolated words, and in noisy environments, comprehension improves once the listener knows the topic of a conversation (Miller et al., 1951; Boothroyd and Nitttrouer, 1988; Grant and Seitz, 2000; Stickney and Assmann, 2001; Davis et al., 2005). Therefore, both intuitive reasoning and objective evidence from these psychoacoustic investigations indicate that speech comprehension is the result of an interaction between bottom-up processes, involving decoding of the speech signal along the auditory pathway, and top-down processes informed by semantic context.

The brain regions that interact to match acoustic information with context are not well understood. Functional imaging studies have advanced our understanding of the functional anatomy of the auditory processes supporting speech perception (Binder et

al., 1996; Scott et al., 2000; Davis and Johnsrude, 2003; Zekveld et al., 2006) (for review, see Scott and Johnsrude, 2003; Poeppel and Hickok, 2004; Xu et al., 2005), with good evidence for an auditory processing stream along the superior temporal gyrus that is sensitive to intelligibility. Noise vocoding is an effective technique to manipulate the spectral detail of speech (Shannon et al., 1995) and render it more or less intelligible in a graded manner (Scott et al., 2000, 2006; Davis and Johnsrude, 2003; Warren et al., 2006). In a previous behavioral experiment, we demonstrated that a contextual manipulation was most effective at an intermediate level of quality of speech signal (J. Obleser, L. Alba-Ferrara, and S. K. Scott, unpublished observation). We used sentences that varied in semantic predictability, so that the strength of semantic associations between the key words was either high or low (e.g., “He caught a fish in his net” vs “Sue discussed the bruise”) (Kalikow et al., 1977; Stickney and Assmann, 2001). Semantic predictability has been shown previously to influence speech perception (Boothroyd and Nitttrouer, 1988; Pichora-Fuller et al., 1995; Stickney and Assmann, 2001), and in our study, we demonstrated that it affected accuracy of comprehension, with an improvement from 50 to 90%, when used with spectrally degraded, noise-vocoded sentences of intermediate degree (eight frequency channels) (Obleser, Alba-Ferrara, and Scott, unpublished observation).

In this study, subjects listened to sentences varying in acoustic degradation and in semantic predictability in an event-related functional magnetic resonance imaging (fMRI) experiment. The aim was to identify the interdependency between bottom-up and top-down processes during speech comprehension, specifically to investigate which brain regions mediate successful yet effortful speech comprehension through contextual information under adverse acoustic conditions and how these brain regions interact.

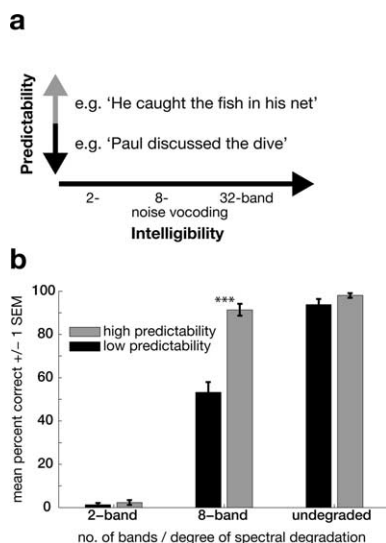
Received Oct. 26, 2006; revised Dec. 18, 2006; accepted Jan. 29, 2007.

This work was supported by the Wellcome Trust (S.K.S.), the Medical Research Council (R.J.S.W.), and the Landesstiftung Baden-Württemberg Germany (J.O.). We are grateful to Lucy Alba-Ferrara for help with acquiring the data and to two anonymous reviewers for their helpful suggestions.

Correspondence should be addressed to Dr. Jonas Obleser, Institute of Cognitive Neuroscience, University College London, 17 Queen Square, London WC1N 3AR, UK. E-mail: j.obleser@ucl.ac.uk.

DOI:10.1523/JNEUROSCI.4663-06.2007

Copyright © 2007 Society for Neuroscience 0270-6474/07/272283-07\$15.00/0



**Figure 1.** *a*, Illustration of the basic experimental design. Predictability (i.e., the inner semantic coherence of a sentence or how well one key word predicts the others) and intelligibility of the signal (through noise vocoding and thereby reducing the spectral detail) are varied orthogonally. *b*, Behavioral pretesting results unequivocally identified predictability to be most effective on speech comprehension at intermediate signal quality of eight-band noise-vocoded speech.  $***p < 0.0001$  [data based on  $n = 18$  (replotted from Obleser, Alba-Ferrara, and Scott, unpublished observation)]. Error bars indicate SEM.

## Materials and Methods

**Subjects.** Sixteen right-handed monolingual speakers of British English (seven females; mean age,  $24 \pm 6$  years SD) were recruited. All were native monolingual speakers of English, and none had a history of a neurological, psychiatric, or hearing disorder. No subject had previous experience of noise-vocoded or spectrally rotated speech, and all were naive to the purpose of the study. The total duration of the procedure was <1 h. The study had previous approval of the local ethics committee of the Hammersmith Hospitals Trust.

**Stimulus material.** The stimulus material consisted of 180 spoken sentences from the SPIN (speech intelligibility in noise) test (Kalikow et al., 1977) (forms 2.1, 2.3, 2.5, and 2.7), half of which comprise low-predictability sentences (e.g., “Sue discussed the bruise” or “They were considering the gang”) and high-predictability sentences (e.g., “His boss made him work like a slave” and “The watchdog gave a warning growl”), matched for phonetic and linguistic variables such as phonemic features, number of syllables, and content words. The sentences were recorded by a phonetically trained female speaker of British English in a soundproof chamber [using a Brüel & Kjær (Naerum, Denmark) 2231 sound-level meter fitted with a 4165 cartridge; sampling rate, 44.1 kHz]. The final set of stimuli was created off-line by down-sampling the audio recordings to 20 kHz (9 kHz bandpass), editing the sentences at zero-crossings before and after each sentence, and applying 5 ms linear fades to onsets and offsets. Sentence recordings were normalized with respect to average root-mean-squared amplitude and had an average duration of 2.2 s.

Each of the 180 sentence recordings was submitted to a noise-vocoding routine (Shannon et al., 1995) with 2, 8, or 32 filter bands, resulting in six conditions (Fig. 1*a*): two predictability levels (low, high) at three intelligibility levels (logarithmically varying spectral degradation through 2-, 8- or 32-band noise vocoding) with 30 stimuli each. A seventh condition was created to serve as an entirely unintelligible control condition: an additional set of 30 noise-vocoded stimuli (32 bands) were presented after they had been spectrally rotated (Blesser, 1972). This control condition has been used in imaging studies (Scott et al., 2000; Narain et al., 2003; Obleser et al., 2007); it leaves the temporal envelope unaltered and preserves the spectrotemporal complexity, whereas the signal is rendered unintelligible by inverting the frequency spectrum.

The exact levels of noise vocoding were chosen after a series of behavioral pretests (Obleser, Alba-Ferrara, and Scott, unpublished observa-

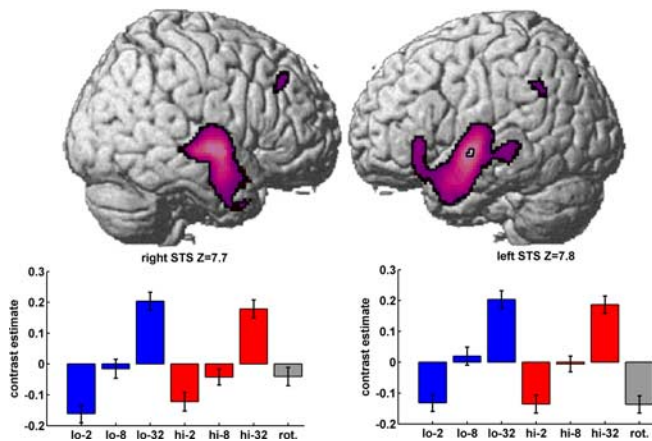
tion). Eight-band noise-vocoded speech had been identified as the condition in which predictability had the largest influence (for a signal frequency range of 0–9 kHz). At such an intermediate signal quality, the influence of context provided by semantic predictability proved most effective in improving performance. In two experiments that closely preserved the design of this imaging study, identification of key words within sentences improved by almost 40% for sentences with high predictability compared with sentences with low predictability. The same sentence material was used and presented randomly, and we orthogonally varied predictability and intelligibility over a wide range of spectral degradation (noise-vocoding) levels. In contrast, there was no influence of predictability on key word recognition either with two-band noise-vocoded speech or with normal speech. In the current imaging study, we approximated normal speech with 32-band noise-vocoded speech to avoid pop-out effects (Fig. 1*b*).

**Experimental procedures.** Subjects were in the supine position in a 3.0T Philips (Best, The Netherlands) Intera scanner equipped with a six-element SENSE head coil for radiofrequency signal detection, fitted with a  $B_0$ -dependent auditory stimulus delivery system (MR-Confon, Magdeburg, Germany). In a short familiarization period, all subjects listened to 30 examples of noise-vocoded speech with three levels of degradation and two levels of predictability. The levels of degradation and the sentences were not used in the subsequent fMRI study. There were also 30 trials on spectrally rotated speech. All subjects recognized the noise-vocoded stimuli as speech, despite the varying degrees of spectral degradation and intelligibility.

In the scanner, subjects were instructed to lie still and listen attentively. They were prepared to answer some questions on the material afterward, while no further task was introduced. A series of 240 MR-volume scans (echo-planar imaging) was obtained. Trials of all seven conditions (30 trials and volumes per condition) and 30 additional silent trials were presented in a pseudo-randomized and interleaved manner. An MR volume consisted of 32 axial slices, obliquely oriented to cover the entire brain (an in-plane resolution of  $2.5 \times 2.5$  mm<sup>2</sup> with a slice thickness of 3.25 mm and a 0.75 mm gap). Scans were acquired with SENSE factor 2 and second-order shim gradients to reduce blurring and signal loss associated with susceptibility gradients adjacent to the ear canals. Volume scans were acquired using temporal sparse sampling (Hall et al., 1999) with a repetition time of 9.0 s, an acquisition time of 2.0 s, and a single stimulus being presented, in silence, 5 s before the next volume scan. The exact stimulus onset time was jittered randomly  $\pm 500$  ms to sample the blood oxygenation level-dependent (BOLD) response more robustly. The functional run lasted for 36 min and was followed by a high-resolution T1-weighted scan to obtain a structural MR image for each subject.

**Data analysis.** Using SPM5 (Wellcome Department of Imaging Neuroscience, London, UK), images were corrected for slice timing, realigned, coregistered, normalized to a T1 template using parameters from gray matter segmentation, and smoothed (8 mm<sup>3</sup> Gaussian kernel). For each subject, seven regressors of interest (seven conditions) and six realignment parameters were modeled using a finite impulse response (Gaab et al., 2006), with the silent trials forming an implicit baseline. At the second (group) level, a random-effects within-subjects ANOVA with seven conditions (each condition contrasted with silence from each subject) was calculated. Unless stated otherwise, all group inferences are reported at an uncorrected level of  $p < 0.005$  and a cluster extent of  $>30$  voxels. Coordinates of peak activations were transformed into Talairach coordinates and labeled according to the Talairach Daemon Database (Lancaster et al., 2000).

**Analysis of functional connectivity.** For clusters of activated voxels in the contrast of high- with low-predictability sentences using eight-band noise-vocoded speech (the level of degradation at which semantic predictability had the greatest behavioral effect) (Fig. 1*a*), a correlation analysis was planned to investigate the strength of functional connectivity between clusters. Using the MarsBaR toolbox within SPM5 and using the first eigenvector to summarize activation of a cluster across voxels, time courses from all significant clusters were extracted from each subject's data across all conditions. The condition- and cluster-specific time courses of all subjects were then collapsed into a median time course that



**Figure 2.** Main effect of intelligibility. Results from an F-contrast for the main effect of intelligibility (number of bands) is shown, yielding extensive bilateral clusters in temporal lobes, with peak voxels (contrast bar graphs for all 7 conditions are shown below) in the left and right STS that are not modulated by predictability. In the left hemisphere, the cluster also incorporates the inferior frontal cortex. The posterior parietal activations show the reverse pattern (decreasing activation with increasing intelligibility). All activations are  $p < 0.005$ ; the cluster extent is  $> 30$  voxels. Error bars indicate SEM. lo, Low; hi, high; rot., rotated.

represents the average activity time course of a given brain region in a given condition (having 30 sampling points for the 30 trials of each condition; because it is averaged across subjects, it also has an enhanced signal-to-noise ratio). Correlations of activity time courses between brain regions were then analyzed separately across all conditions and assessed statistically using Pearson’s correlation coefficient.

**Results**

We found extensive bilateral temporal activation in response to all stimuli in all subjects. Therefore, all subjects were included in the random-effects group analysis. First, an F-contrast for any main effect of intelligibility was assessed (i.e., we looked for brain regions that showed a change in activity with increasing spectral detail in the speech signal, regardless of predictability). This analysis revealed extensive bilateral activation in the temporal lobes, with the peak voxels in each hemisphere located in the anterior superior temporal sulcus (STS), with extension on the left into the inferior frontal gyrus (Fig. 2). All of these regions showed a quasi-monotonic increase in BOLD signal with increasing intelligibility of the signal. In contrast, the medial parietal cortex (precuneus) and the left posterior inferior parietal cortex demonstrated decreasing activation with increasing signal quality [see Table 1 for the stereotactic coordinates, in MNI (Montreal Neurological Institute) space, for the peak voxels].

Second, the influence of sentence predictability at an intermediate degradation level was investigated, informed by the known behavioral effect at intermediate signal quality (eight-band noise-vocoded speech), the condition when predictability had the greatest influence on speech comprehension (Fig. 1). The left and right anterior STS showed no difference in activity between low- and high-predictability eight-band noise-vocoded sentences. However, activity with high-predictability sentences extended posteriorly into the left posterior temporal and inferior parietal cortices and forward into the left temporal pole and ventral inferior frontal cortex (Fig. 3). There were additional activations outside the temporal lobes, in the medial prefrontal and posterior cingulate cortices.

A direct comparison of brain activity in response to sentences of high and low predictability with the eight-band noise-vocoded sentences confirmed the activation of these cortical areas (Fig. 3,

Table 1). Five brain regions, four lateralized to the left hemisphere and one midline in the anterior prefrontal cortex, demonstrated increased activity in response to degraded yet highly predictable speech. The left dorsolateral prefrontal cortex, angular gyrus, and posterior cingulate cortex did so only under this condition of effortful yet successful speech comprehension. Importantly, activity returned to baseline when the sentences were both highly predictable and readily intelligible; therefore, these regions were not demonstrating a response simply to success at comprehension, although this effect was observed in the medial prefrontal cortex and in the left inferior frontal gyrus, where activity did not differ between 8- and 32-band noise-vocoded speech when sentences were of high predictability.

Third, the functional connectivity between these five cortical areas was determined, as described in Materials and Methods. Because these cortical areas were engaged when the degraded speech signal was heard within the context of high predictability (i.e., when comprehension was enhanced by semantic context), a change in functional connectivity was expected (expressed as an across-trials correlation between one cortical area and another). This prediction was confirmed as an increase in correlation of the activity between cortical areas when the eight-band noise-vocoded sentences were of high predictability (Table 2). Notably, the correlation of the responses between the left angular gyrus and prefrontal cortex changed from being not significant ( $r = 0.12$  and  $r = 0.25$  for the lateral and medial prefrontal cortex, respectively) when the sentences were of low predictability to significant ( $r = .68$  and  $r = .71$ , respectively;  $p < 0.0001$ ) when sentences were of high predictability (Fig. 4). Although the hypothesis-led analyses were directed at the comparison of sentences of high versus low predictability heard at an intermediate signal quality, we also observed that the correlation of activity between brain regions was not significant when subjects heard unintelligible rotated speech. The correlations between activated areas within the prefrontal cortex and between anterior and posterior midline areas were high, regardless of whether the eight-band noise-vocoded sentences were of high or low predictability (i.e., predictability did not modulate the functional connectivity between these areas) (Table 2).

**Discussion**

We have demonstrated how changes in functional integration across very distributed brain regions improve speech perception under acoustically suboptimal conditions. Using a design that varied orthogonally intelligibility (Shannon et al., 1995) and semantic predictability (Stickney and Assmann, 2001) revealed functional connections between areas in the temporal, inferior parietal, and prefrontal cortices that strengthened and supported comprehension of sentences with high semantic predictability but intermediate signal quality. Because the signal quality was constant (eight-band noise-vocoded speech) across sentences of both low and high predictability, this effect could be attributed to the modulating effect of semantic context (Kalikow et al., 1977; Stickney and Assmann, 2001) (Obleser, Alba-Ferrara, and Scott, unpublished observation). This was further confirmed by the observation that the strengthening of functional connections was between areas that were not responding simply to increased sentence comprehension, because activity was not maintained in response to the easily understood 32-band noise-vocoded sentences.

Activation in the contrast of high- and low-predictable eight-band noise-vocoded sentences was most evident in the left angular gyrus. Activity in this area, and at the other extreme of the left

temporal lobe, in the anterior temporal cortex extending into the ventral inferior frontal gyrus, showed this effect of high predictability more than low predictability at intermediate signal quality. Additional areas demonstrating the same effect were the left dorsolateral and medial prefrontal cortices and the left posterior cingulate cortex. In contrast, the lateral temporal neocortex within the superior temporal gyrus and sulcus was sensitive to the increasing spectral detail across all stimuli.

The enhanced activity was seen only in the high-predictability eight-band condition; when speech comprehension was effortless in response to 32-band speech, activity in the angular gyrus, posterior cingulate cortex, and dorsolateral prefrontal cortex returned to baseline level (Fig. 3). In other words, only if speech comprehension succeeds despite adverse acoustic conditions are these regions involved. Changes in predictability in the absence of signal degradation (i.e., for 32-band signals) were not accompanied by substantial increases in brain activation, nor did they lead to differences in speech recognition in the behavioral pretests. The conclusion is that the influence of semantic context when listening to short sentences only becomes crucial once the signal is compromised.

Functional connectivity analysis among the activated clusters yielded positive correlations between their time courses of activity. This evidence for functional integration was most evident when the subjects listened to eight-band noise-vocoded speech that could be decoded because of semantic context compared with the corresponding signal when semantic context was absent. As summarized in Table 2, the connectivity between the angular gyrus and the other four activation clusters showed the greatest increase (Fig. 4). Interestingly, this strength of connectivity between the angular gyrus and the frontal lobe is reduced in developmental dyslexia (Horwitz et al., 1998), and a recent study demonstrated parietal-frontal “underconnectivity” during sentence comprehension in autistic subjects (Kana et al., 2006). The strengthened connectivity along the temporal lobe (between the angular gyrus and temporal pole) fits well with recent evidence on anatomical links between these areas (middle longitudinal fasciculus), as does the link between the angular gyrus and lateral prefrontal cortex (Schmahmann and Pandya, 2006). A recent study on written text comprehension also identified an increase in activity in the angular gyrus [Brodmann’s area (BA) 39] in a contrast of real-word sentences with sentences comprising pseudowords (Ferstl and von Cramon, 2002). This is additional evidence that the angular gyrus is a resource for semantic processing, activated in our study when semantics had a decisive influence on speech perception.

Because all of the regions are distributed across the frontal and parietal cortices, their contribution to speech comprehension is likely to be of higher order than basic acoustic processing. These widespread activations are likely to represent a number of cognitive-supporting mechanisms, among them aspects of working memory and attention. One hypothesis is that the contribution of the angular gyrus is through its role in verbal working memory, and it is a region that has frequently been implicated in explicit semantic decision tasks (Binder et al., 2003; Scott et al., 2003; Sharp et al., 2004). Other processes such as phonological memory and auditory–motor transformation processes (Jacque-

**Table 1. Overview of significant clusters in random-effects analysis ( $p < 0.005$ ; cluster extent,  $> 30$  voxels)**

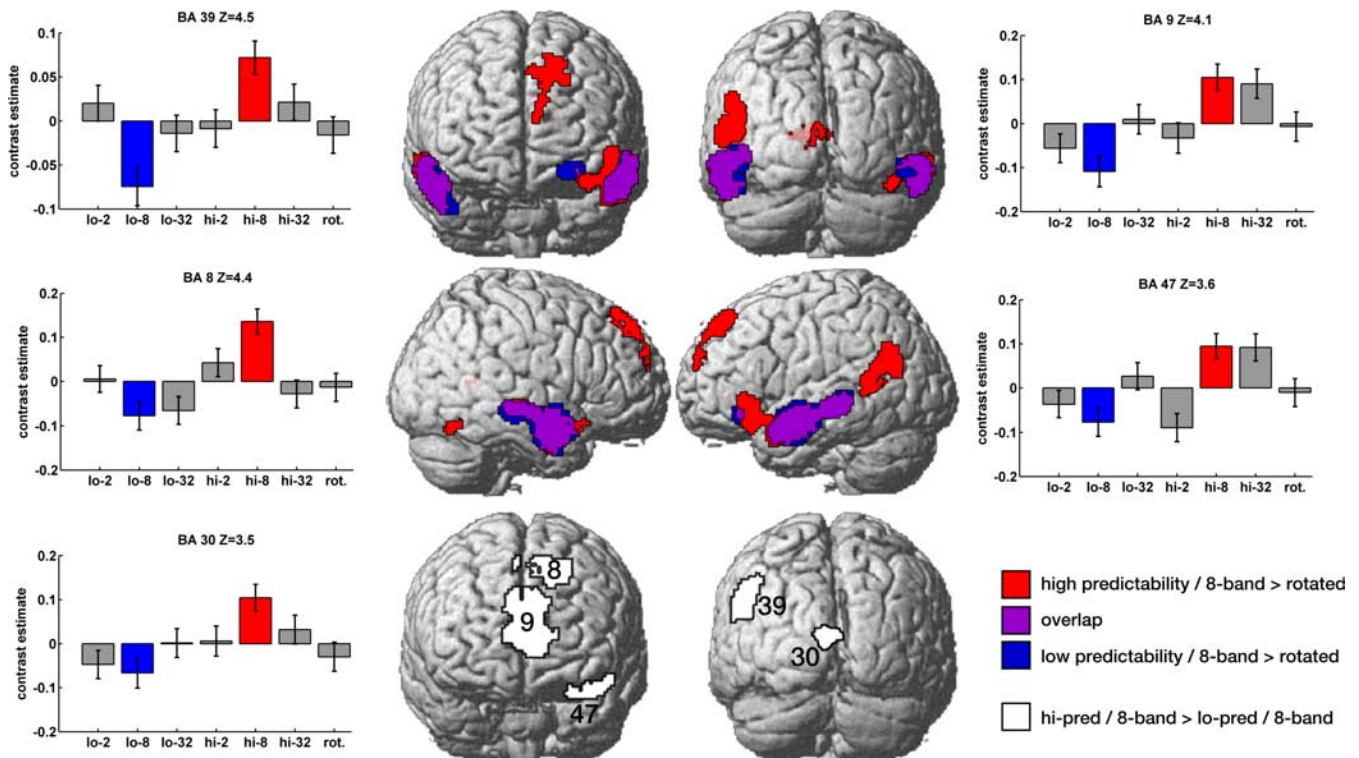
Site	MNI coordinates	Z	Extent (mm <sup>3</sup> )
<b>Main effect of intelligibility</b>			
L. STS (BA 21)	−57, −6, −5	7.79	19,848
R. STS (BA 21)	59, −6, −3	7.73	14,048
L. precuneus (BA 7)	−14, −72, 37 −38, −14,	3.59	824
L. middle temporal gyrus (BA 21)	−11	3.39	432
R. middle frontal gyrus (BA 9)	30, 37, 37	3.36	536
L. inferior parietal lobe (BA 40)	−55, −51, 36	3.19	336
R. precuneus (BA 7)	8, −62, 40	3.07	752
<b>High-predictability 8-band speech &gt; low-predictability 8-band speech</b>			
L. angular gyrus (BA 39)	−50, −60, 34	4.50	2320
L. superior frontal gyrus (BA 8)	−20, 39, 44	4.39	2656
Medial superior frontal gyrus (BA 9)	4, 56, 27,	4.10	10,024
L. ventral inferior frontal gyrus (BA 47)	−40, 24, −18	3.59	1184
L. posterior cingulate cortex (BA 29/30)	−4, −56, 12	3.38	912

Specifications refer to peak voxels. MNI, Montreal Neurological Institute; L, left; R, right.

mot et al., 2003; Hickok and Poeppel, 2004; Warren et al., 2005; Jacquemot and Scott, 2006) that might contribute to recovering meaning from degraded speech signals are located in adjacent but separate temporo-parietal areas. Working memory operations in the left dorsolateral prefrontal cortex [Petrides, 1994; Owen, 1997; for a review on working memory, see Owen et al. (2005)] might support comprehension under adverse conditions by permitting the manipulation of the degraded stimuli within short-term memory. Thus, components of the degraded sentence that do not map automatically onto meaning can be reconstructed by reprocessing them within the context of semantic predictability.

The fronto-parietal network seen here might also reflect monitoring and selection processes more commonly associated with attention than only maintaining information in short-term memory (Lebedev et al., 2004). This interpretation postulates that the prefrontal cortex has a role in directing attention to relevant auditory features, to guide both short-term memory and access to long-term memory representations (for review, see Miller and Cohen, 2001). Most relevant to comprehending distorted speech in a sentence context is the concept of competition among lexical and phonological candidates, because signal degradation introduces considerable ambiguity. A recently suggested framework for the prefrontal cortex and conceptual selection problems (Kan and Thompson-Schill, 2004) would imply that the system for speech perception has to solve problems associated with lexical selection. Because of the acoustic ambiguity, each key word might activate multiple possible word candidates. With high semantic coherence, top-down influences guide correct lexical selection, but with low semantic coherence, lexical selection will be much less successful. For such selection and competition processes, the left prefrontal cortex is engaged (Tippett et al., 2004). Also, top-down control of selective attention in our study might encompass the on-line formation of increasingly specific hypotheses about which sentence-final word to expect in a degraded yet predictable sentence. This in turn would enable more thorough and, ultimately, more successful (re-)analysis of the noise-vocoded signal as more elements of the sentence become available.

Finally, the facilitation through context in the current stimulus set is likely to entail a range of possible subordinate lexical mechanisms by which predictability supports comprehension. Both verb semantics (by narrowing down the context) and semantic associations in general (by allowing to “prime” for other word candidates) are possible influences here (Kalikow et al.,



**Figure 3.** Overview over differential activation for high- versus low-predictability sentences at intermediate (8-band) signal quality. The top and middle panels of activation overlays display the two conditions separately (in red and blue, respectively) compared with spectrally rotated speech. The bottom panels of brain overlays show the direct comparison of both. Bar graphs show activations in cluster peak voxels of direct comparison. BA 39, BA 8, and BA 30 (left panels) and BA 9 and BA 47 (right panels) all exhibit strongest activation through high-predictability eight-band speech. All activations are  $p < 0.005$ ; the cluster extent is  $> 30$  voxels. Error bars indicate SEM. lo, Low; hi, high; rot., rotated; hi-pred, high predictability; lo-pred, low predictability.

**Table 2. Changes in correlation comparing high-predictability and low-predictability eight-band conditions ( $\Delta r$ )**

	BA 39	BA 8	BA 9	BA 47	BA 30
BA 39		0.5553 (0.12)	0.4560 (0.25)	0.2393 (0.36 <sup>+</sup> )	0.2344 (0.20)
Lo-pred. 8-band					
Hi-pred. 8-band		(0.68 <sup>***</sup> )	(0.71 <sup>***</sup> )	(0.60 <sup>***</sup> )	(0.43 <sup>*</sup> )
BA 8			0.1155 (0.69 <sup>***</sup> ) (0.81 <sup>***</sup> )	0.2243 (0.52 <sup>**</sup> ) (0.75 <sup>***</sup> )	-0.0550 (0.55 <sup>***</sup> ) (0.50 <sup>**</sup> )
BA 9				-0.0126 (0.80 <sup>***</sup> ) (0.78 <sup>***</sup> )	-0.0795 (0.62 <sup>***</sup> ) (0.54 <sup>**</sup> )
BA 47					0.0291 (0.51 <sup>**</sup> ) (0.54 <sup>**</sup> )
BA 30					

Positive values indicate an increase in positive correlation between brain regions in the high-predictability eight-band condition compared with the low-predictability eight-band condition. <sup>+</sup> $p < 0.10$ ; <sup>\*</sup> $p < 0.05$ ; <sup>\*\*</sup> $p < 0.01$ ; <sup>\*\*\*</sup> $p < 0.001$ . Correlations of  $r > 0.50$  would be considered significant on Bonferroni correction ( $n = 16$ ; based on the median of within-subject eigenvector time courses from activation clusters;  $p < 0.005$ ;  $k > 30$ ; random-effects). Lo-pred., Low predictability; Hi-pred., high predictability.

1977; Friederici and Kotz, 2003), a matter to disentangle in additional studies.

To summarize, successful speech perception under less than ideal listening conditions depends on greater functional integration across a very distributed left-hemispheric network of cortical areas. Therefore, speech perception is facilitated when high-order

cognitive subsystems become engaged, and it cannot be considered as the product of processing within the unimodal auditory cortex alone (Jacquemot and Scott, 2006).

A clear lateralization to the left was observed. Left-hemisphere predominance is often absent from studies of speech perception and comprehension, and it was by and large absent in the main effect of intelligibility results in the present study (Fig. 2). Thus, it appears that only once higher-order processes interact with downstream perceptual systems is the left lateralization established. Notably, left- and right-hemispheric temporal cortices also varied in their degree of responsiveness to the entirely unintelligible sounds of rotated speech: as can be seen in Figure 2, right STS areas show near-baseline activation in response to rotated speech, whereas the left STS area clearly exhibits a relative deactivation [compare previous findings on rotated speech (Scott et al., 2000; Narain et al., 2003)].

Interestingly, Broca's area (BA 44/45) was not activated differentially by high and low predictability despite its known role in rule-based processing also in speech. Its response pattern was statistically indistinguishable from the anterolateral STS (Fig. 2). Thus, structural processing of the language content in our set of stimuli may only have become possible once a certain level of signal quality was attained, and that it increased further as perceptual ambiguity was overcome.

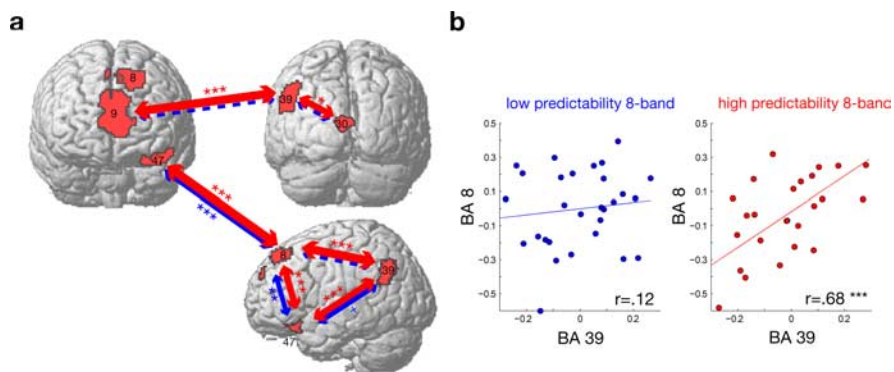
By using a factorial design, parametric degradation of the speech signal and an acoustically matched unintelligible baseline, we have not only been able to confirm the network of cortical areas that respond to speech intelligibility (Binder et al., 2000; Scott et al., 2000, 2004; Davis and Johnsrude, 2003; Obleser et al., 2006; Zekveld et al., 2006) but have demonstrated the dynamic

changes within the network that occur when intelligibility and semantic context interact. This captures the nature of speech perception under real-life conditions.

In conclusion, the combination of two established manipulations of speech, one acoustic (spectral degradation through noise vocoding) and the other linguistic (semantic predictability), allowed us to demonstrate a widely distributed left-hemisphere array of cortical areas that can establish speech perception under adverse listening conditions. These areas are remote from the unimodal auditory cortex in high-order heteromodal and amodal cortices. Their functional connectivity is strengthened when semantic context has a beneficial influence on speech perception.

## References

- Binder JR, Frost JA, Hammeke TA, Rao SM, Cox RW (1996) Function of the left planum temporale in auditory and linguistic processing. *Brain* 119:1239–1247.
- Binder JR, Frost JA, Hammeke TA, Bellgowan PS, Springer JA, Kaufman JN, Possing ET (2000) Human temporal lobe activation by speech and non-speech sounds. *Cereb Cortex* 10:512–528.
- Binder JR, McKiernan KA, Parsons ME, Westbury CF, Possing ET, Kaufman JN, Buchanan L (2003) Neural correlates of lexical access during visual word recognition. *J Cogn Neurosci* 15:372–393.
- Blessner B (1972) Speech perception under conditions of spectral transformation. I. Phonetic characteristics. *J Speech Hear Res* 15:5–41.
- Boothroyd A, Nittroter S (1988) Mathematical treatment of context effects in phoneme and word recognition. *J Acoust Soc Am* 84:101–114.
- Davis MH, Johnsrude IS (2003) Hierarchical processing in spoken language comprehension. *J Neurosci* 23:3423–3431.
- Davis MH, Johnsrude IS, Hervais-Adelman A, Taylor K, McGettigan C (2005) Lexical information drives perceptual learning of distorted speech: evidence from the comprehension of noise-vocoded sentences. *J Exp Psychol Gen* 134:222–241.
- Ferstl EC, von Cramon DY (2002) What does the frontomedian cortex contribute to language processing: coherence or theory of mind? *NeuroImage* 17:1599–1612.
- Friederici AD, Kotz SA (2003) The brain basis of syntactic processes: functional imaging and lesion studies. *NeuroImage* 20 [Suppl 1r]:S8–S17.
- Gaab N, Gabrieli JDE, Glover GH (2007) Assessing the influence of scanner background noise on auditory processing I. An fMRI study comparing three experimental designs with varying degrees of scanner noise. *Hum Brain Mapp*, in press.
- Grant KW, Seitz PF (2000) The recognition of isolated words and words in sentences: individual variability in the use of sentence context. *J Acoust Soc Am* 107:1000–1011.
- Hall DA, Haggard MP, Akeroyd MA, Palmer AR, Summerfield AQ, Elliott MR, Gurney EM, Bowtell RW (1999) “Sparse” temporal sampling in auditory fMRI. *Hum Brain Mapp* 7:213–223.
- Hickok G, Poeppel D (2004) Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition* 92:67–99.
- Horwitz B, Rumsey JM, Donohue BC (1998) Functional connectivity of the angular gyrus in normal reading and dyslexia. *Proc Natl Acad Sci USA* 95:8939–8944.
- Jacquemot C, Scott SK (2006) What is the relationship between phonological short-term memory and speech processing? *Trends Cogn Sci* 10:480–486.
- Jacquemot C, Pallier C, LeBihan D, Dehaene S, Dupoux E (2003) Phonological grammar shapes the auditory cortex: a functional magnetic resonance imaging study. *J Neurosci* 23:9541–9546.
- Kalilow DN, Stevens KN, Elliott LL (1977) Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *J Acoust Soc Am* 61:1337–1351.



**Figure 4.** *a*, Overview over the changes in correlation between brain regions for the high-predictability eight-band condition (red) and the low-predictability eight-band condition (blue). Numbers in clusters indicate BAs of cluster peak voxels. All arrows shown indicate an increase in the positive correlation for the high-predictability condition, whereas no brain areas are more strongly interlinked for low predictability.  $^+p < 0.10$ ;  $*p < 0.05$ ;  $**p < 0.01$ ;  $***p < 0.001$ . The dashed blue lines indicate nonsignificant correlations. For details of changes in correlation ( $\Delta r$ ), see Table 2. *b*, Scatterplots with linear regression slopes exemplifying the increase in the positive correlation between the left posterior temporal/inferior parietal cortex (angular gyrus, BA 39; x-axis) and left dorsolateral prefrontal cortex (superior frontal gyrus, BA 8; y-axis). Values plotted are condition-specific first eigenvector time series from the cluster identified in random-effects group analysis averaged across subjects (see Materials and Methods).

- Kan IP, Thompson-Schill SL (2004) Selection from perceptual and conceptual representations. *Cogn Affect Behav Neurosci* 4:466–482.
- Kana RK, Keller TA, Cherkassky VL, Minshew NJ, Just MA (2006) Sentence comprehension in autism: thinking in pictures with decreased functional connectivity. *Brain* 129:2484–2493.
- Lancaster JL, Woldorff MG, Parsons LM, Liotti M, Freitas CS, Rainey L, Kochunov PV, Nickerson D, Mikiten SA, Fox PT (2000) Automated Talairach atlas labels for functional brain mapping. *Hum Brain Mapp* 10:120–131.
- Lebedev MA, Messinger A, Kralik JD, Wise SP (2004) Representation of attended versus remembered locations in prefrontal cortex. *PLoS Biol* 2:e365.
- Miller EK, Cohen JD (2001) An integrative theory of prefrontal cortex function. *Annu Rev Neurosci* 24:167–202.
- Miller GA, Heise GA, Lichten W (1951) The intelligibility of speech as a function of the context of the test materials. *J Exp Psychol* 41:329–335.
- Narain C, Scott SK, Wise RJ, Rosen S, Leff A, Iversen PM (2003) Defining a left-lateralized response specific to intelligible speech using fMRI. *Cereb Cortex* 13:1362–1368.
- Obleser J, Boecker H, Drzezga A, Haslinger B, Hennenlotter A, Roettinger M, Eulitz C, Rauschecker JP (2006) Vowel sound extraction in anterior superior temporal cortex. *Hum Brain Mapp* 27:562–571.
- Obleser J, Zimmermann J, Van Meter JW, Rauschecker JP (2007) Multiple stages of auditory speech perception reflected in event-related fMRI. *Cereb Cortex*, in press.
- Owen AM (1997) The functional organization of working memory processes within human lateral frontal cortex: the contribution of functional neuroimaging. *Eur J Neurosci* 9:1329–1339.
- Owen AM, McMillan KM, Laird AR, Bullmore E (2005) N-back working memory paradigm: a meta-analysis of normative functional neuroimaging studies. *Hum Brain Mapp* 25:46–59.
- Petrides M (1994) Frontal lobes and behaviour. *Curr Opin Neurobiol* 4:207–211.
- Pichora-Fuller MK, Schneider BA, Daneman M (1995) How young and old adults listen to and remember speech in noise. *J Acoust Soc Am* 97:593–608.
- Poeppel D, Hickok G (2004) Towards a new functional anatomy of language. *Cognition* 92:1–12.
- Schmahmann JD, Pandya DN (2006) Fiber pathways of the brain. Oxford: Oxford UP.
- Scott SK, Johnsrude IS (2003) The neuroanatomical and functional organization of speech perception. *Trends Neurosci* 26:100–107.
- Scott SK, Blank CC, Rosen S, Wise RJ (2000) Identification of a pathway for intelligible speech in the left temporal lobe. *Brain* 123:2400–2406.
- Scott SK, Leff AP, Wise RJ (2003) Going beyond the information given: a

- neural system supporting semantic interpretation. *NeuroImage* 19:870–876.
- Scott SK, Rosen S, Wickham L, Wise RJ (2004) A positron emission tomography study of the neural basis of informational and energetic masking effects in speech perception. *J Acoust Soc Am* 115:813–821.
- Scott SK, Rosen S, Lang H, Wise RJ (2006) Neural correlates of intelligibility in speech investigated with noise vocoded speech—a positron emission tomography study. *J Acoust Soc Am* 120:1075–1083.
- Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M (1995) Speech recognition with primarily temporal cues. *Science* 270:303–304.
- Sharp DJ, Scott SK, Wise RJ (2004) Monitoring and the controlled processing of meaning: distinct prefrontal systems. *Cereb Cortex* 14:1–10.
- Stickney GS, Assmann PF (2001) Acoustic and linguistic factors in the perception of bandpass-filtered speech. *J Acoust Soc Am* 109:1157–1165.
- Tippett LJ, Gendall A, Farah MJ, Thompson-Schill SL (2004) Selection ability in Alzheimer's disease: investigation of a component of semantic processing. *Neuropsychology* 18:163–173.
- Warren JD, Scott SK, Price CJ, Griffiths TD (2006) Human brain mechanisms for the early analysis of voices. *NeuroImage* 31:1389–1397.
- Warren JE, Wise RJ, Warren JD (2005) Sounds do-able: auditory-motor transformations and the posterior temporal plane. *Trends Neurosci* 28:636–643.
- Xu J, Kemeny S, Park G, Frattali C, Braun A (2005) Language in context: emergent features of word, sentence, and narrative comprehension. *NeuroImage* 25:1002–1015.
- Zekveld AA, Heslenfeld DJ, Festen JM, Schoonhoven R (2006) Top-down and bottom-up processes in speech comprehension. *NeuroImage* 32:1826–1836.