The neural processing of masked speech: Evidence for different mechanisms in the left and right temporal lobes

Sophie K. Scott[a)]
Institute of Cognitive Neuroscience, University College London, 17 Queen Square, London WC1N 3AR, United Kingdom

Stuart Rosen
Division of Psychology and Language Sciences, University College London, London WC1E 6BT, United Kingdom

C. Philip Beaman and Josh P. Davis
Department of Psychology, University of Reading, Whiteknights, Reading RG6 6AL, United Kingdom

Richard J. S. Wise
MRC Clinical Sciences Centre, London W12 0NN, United Kingdom

(Received 5 November 2007; revised 3 November 2008; accepted 13 November 2008)

It has been previously demonstrated that extensive activation in the dorsolateral temporal lobes associated with masking a speech target with a speech masker, consistent with the hypothesis that competition for central auditory processes is an important factor in informational masking. Here, masking from speech and two additional maskers derived from the original speech were investigated. One of these is spectrally rotated speech, which is unintelligible and has a similar (inverted) spectrotemporal profile to speech. The authors also controlled for the possibility of "glimpsing" of the target signal during modulated masking sounds by using speech-modulated noise as a masker in a baseline condition. Functional imaging results reveal that masking speech with speech leads to bilateral superior temporal gyrus (STG) activation relative to a speech-in-noise baseline, while masking speech with spectrally rotated speech leads solely to right STG activation relative to the baseline. This result is discussed in terms of hemispheric asymmetries for speech perception, and interpreted as showing that masking effects can arise through two parallel neural systems, in the left and right temporal lobes. This has implications for the competition for resources caused by speech and rotated speech maskers, and may illuminate some of the mechanisms involved in informational masking. *© 2009 Acoustical Society of America.* [DOI: 10.1121/1.3050255]

PACS number(s): 43.71.Rt, 43.71.Qr [RYL] Pages: 1737–1743

I. INTRODUCTION

The properties of masking sounds affect the extent to which they compete for the same resources—central or peripheral—as the target. The aspects of these properties can be very broadly captured by the terms informational and energetic masking, where in the latter the effects are largely due to competition at the auditory periphery, whereas in the former competition for resources seems to be associated with more central auditory processes. For any particular masking signal, the masking effects typically arise from a combination of energetic and informational factors. For example, while masking speech with steady-state noise will presumably be dominated by energetic effects, masking speech with speech will involve both energetic and informational masking. In this paper we used functional neuroimaging to contrast two different speech-related masking signals. Our aim was to identify any difference in their effects in cortical processing, which could be linked to competition for central auditory processing resources, and thus to aspects of informational masking.

We have previously presented data from a positron emission tomography (PET) study indicating that neural correlates of the functional differences between informational and energetic masking can be distinguished (Scott *et al.*, 2004). Subjects were instructed to listen to a single talker in the presence of either a concurrent, continuous masking noise (energetic masking) or speech from another talker (energetic plus informational masking). Each masker type was presented at four different signal-to-noise ratios (SNRs). For the noise masker, there were level dependent effects in the left ventral prefrontal cortex and supplementary motor area, and level independent effects in the left prefrontal and right posterior parietal cortex. For the speech masker, in contrast, there was a level independent activation extensively through auditory association cortex in regions, lateral, anterior, and posterior to primary auditory cortex. In the left hemisphere, these regions have been previously demonstrated to be important for the early perceptual processing of speech (Jacquemot *et al.*, 2003; Scott and Johnsrude, 2003; Wise *et al.*, 2001), and in the right hemisphere these regions have been

[a)] Author to whom correspondence should be addressed. Electronic mail: sophie.scott@ucl.ac.uk associated with nonverbal aspects of speech perception (Scott *et al.*, 2000; Patterson *et al.*, 2002). We interpreted such activation as evidence implicating neural systems important in speech processing when speech is masked by speech—perhaps due to competition for perceptual resources, or because some of these regions are important in the representation of multiple sources of acoustic information (Zatorre *et al.*, 2004).

A limitation of our previous study was that unmodulated noise with the same spectrum as speech was used as the energetic masking condition. Continuous noise was selected as the energetic masker as it leads to the greatest levels of masking, but this did mean that neural activation in the speech masking conditions associated with "glimpses" of the target and masking speech could not be distinguished from processes more strongly linked to informational masking generally (Festen and Plomp, 1990). Thus, some of the results seen in auditory cortical fields could have been associated with essentially energetic processes by allowing glimpses of the target.

A second limitation of this study is that the precise nature of the speech masking effects is hard to determine—we are unable to distinguish between the effects of the acoustical or lexical properties of the masking speech. Although it can be hard to specifically draw a line between informational and energetic masking effects, it has been established that the maximal informational masking is achieved when masking a talker with the same talker, which indicates an important role for acoustic properties. There is also some role for lexical information in speech masking effects (Brungart, 2001), since intrusions from masking speech occur at rates higher than those expected by chance.

Either of these mechanisms (acoustic or linguistic processing), as well as glimpses (which are naturally acoustic in nature), could be responsible for the activation seen in our previous study. Functional imaging studies are well positioned to be able to determine the contributions of these different factors to masking by speech. The bilateral temporal lobe systems recruited by speech perception can be fractionated, both in terms of hemispheric asymmetries and along anatomical lines. Functional imaging studies have shown a clear dominance for left superior temporal areas in the processing of linguistic information in speech (Scott *et al.*, 2000; Narain *et al.*, 2003; Jacquemot *et al.*, 2003). In contrast, right superior temporal areas consistently respond to signals with pitch variation, be these in speech or music (Patterson *et al.*, 2002; Scott *et al.*, 2000; Zatorre and Belin 2001). Functional imaging thus has the power to differentiate linguistic from nonlinguistic processing of masking speech. The aim of the current study is to identify the way in which masking speech competes for central auditory processes, and the extent to which this relates to linguistic processes, and to attempt to control for the possibility of glimpses contributing to the effects previously reported.

Several behavioral papers have interrogated aspects of informational masking by using speech and time-reversed speech as maskers (Hawley *et al.*, 2004; Rhebergen *et al.*, 2005; Johnstone and Litovsky, 2006). In this study we use spectrally rotated speech as a comparison masker (Blesser, 1972), since it has several advantages over reversed speech in terms of its acoustic profile (Scott and Wise, 2004). Hence, three different stimuli were used as maskers: speech, rotated speech (Blesser, 1972), and noise with the same long-term spectrum as speech, modulated by the envelope of speech [speech-modulated noise (SMN); Festen and Plomp, 1990]. These stimuli have different acoustic and lexical characteristics, and imaging the processing of these maskers will give some indication about which characteristics are processed in masking signals. Furthermore, we will be able to establish whether the neural systems responsible for processing characteristics of maskers are similar to those already implicated in cortical speech processing (e.g., Scott *et al.*, 2000; Mummery *et al.*, 1999).

The first two maskers—speech and rotated speech—have very similar auditory profiles, although only the speech also has lexical information. Rotated speech has the spectrotemporal complexity of speech, and maintains much of the sense of voice pitch variation, but is unintelligible. Masking from rotated speech would therefore be associated with prelexical, acoustic aspects of the signal. SMN has the same amplitude modulations as the original speech signal but none of the spectral complexity, structure, or sense of pitch. As a masker it thus allows glimpses of the target speech. The use of this as a baseline masking condition allows us not only to contrast speech masking conditions with noise masking conditions but to also control for the possibility of glimpses, periods during which the masker energy is relatively low, so that the target speech is more readily heard (Festen and Plomp, 1990). This ensures that activation detected when contrasting speech-in-speech over speech-in-noise does not arise simply from "glimpsing" during amplitude "dips."

We have two main hypotheses. By contrasting speech-in-speech and speech-in-rotated-speech with speech-in-SMN, we are controlling for glimpses of the target stimuli. If we see cortical activation associated with these speech based maskers, therefore, we can conclude that this is associated with central auditory processing of the masking signal. Our second hypothesis is that there will be differences in the central processing of the speech and rotated speech maskers, with speech being processed bilaterally (as it contains both acoustic and linguistic elements of speech) while rotated speech will be associated with right temporal lobe activation (as it does not contain the linguistic elements of speech).

II. METHODS: STIMULUS PREPARATION

Three different sets of stimuli were constructed: speech-in-speech, speech-in-rotated-speech, and speech-in-SMN. Oscillograms and spectrograms for each masker type are shown in Fig. 1. All stimulus materials were drawn from digital representations (sampled originally at 44.1 kHz) of simple sentences recorded in an anechoic chamber by a male and a female talker of standard Southern British English. The target sentences were always Bamford–Kowal–Bench (BKB) sentences (Foster *et al.*, 1993) spoken by a female whereas maskers were based on the Institute of Hearing Research audio-visual sentences lists spoken by a male (MacLeod and Summerfield, 1987). All sentences were low-pass filtered at

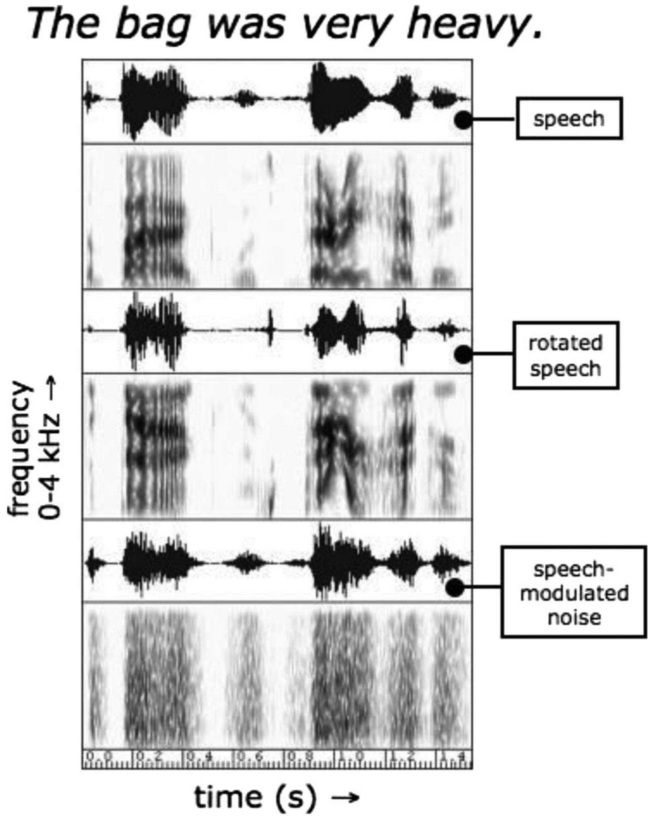

FIG. 1. Oscillograms and spectrograms for the three masking stimuli: speech, rotated speech, and SMN.

3.8 kHz (6th-order elliptical filter, both forward and backward, so as to ensure zero-phase filtering equivalent to a 12th-order filter), and then downsampled to 11.025 kHz to save space. For the speech masker conditions, the masker sentences underwent no further processing. Rotated speech maskers were spectrally inverted using a digital version of the simple modulation technique described by Blesser (1972). In order to preserve somewhat more of the high frequency energy in the original speech signal, here the signals were inverted around 2 kHz, instead of the 1.6 kHz used by Blesser (1972). Because normal and spectrally inverted signals would lead to sounds with very different long-term spectra, the speech signal was first equalized with a filter (essentially high pass) that would make the inverted signal have approximately the same long-term spectrum as the original. This equalizing filter was constructed on the basis of recent extensive measurements of the long-term average spectrum (LTAS) of speech (Byrne *et al.*, 1994), and implemented in finite impulse response (FIR) form. The equalized signal was then amplitude-modulated by a sinusoid at 4 kHz, followed by forward-backward low-pass filtering at 3.8 kHz as described above. The total rms level of the inverted signal was set equal to that of the original low-pass filtered signal.

SMN was created by modulating a speech-shaped noise with envelopes extracted from the original wide-band masker speech signal by full-wave rectification and second-order Butterworth low-pass filtering at 20 Hz. The speech-shaped noise was based on a smoothed version of the LTAS of the male masker sentences. All 270 masker sentences (sampled at 22.05 kHz) were subjected to a spectral analysis using a fast Fourier transform (FFT) of length 512 sample points (23.22 ms), with windows overlapping by 256 points, giving a value for the LTAS at multiples of 43.1 Hz. This spectrum was then smoothed (in the frequency domain) with a 27-point Hamming window that was two-octaves wide, over the frequency range 50 Hz–7 kHz. The smoothed spectrum was then used to construct an amplitude spectrum for an inverse FFT (assuming a sampling rate of 11.025 kHz) with component phases randomized with a uniform distribution over the range $0-2\pi$.

Sentences at different SNRs were created by digital addition, with SNRs determined by a simple rms calculation across the entire waveform. All combined waves were normalized to the same rms value. Because sentences were typically of different durations, summation of the original sentences would have meant that either the target or the masker would have sound energy at its end occurring in a period of silence of its pair (assuming onsets were synchronized). Sentence pairs were thus modified in duration to their mean using the synchronized overlap-and-add (SOLA) technique (Roucus and Wilgus, 1985) as implemented by Huckvale (2007). This alters the duration of speech without changing its fundamental frequency or spectral properties. SOLA cannot, in fact, guarantee any particular final duration, but analysis of the sentences after processing showed them all to fall within a 15 ms range (around a mean of 1.545 s). The shorter sentence in each pair was padded with an appropriate number of zeros before the final addition.

III. BEHAVIORAL TESTING

The intelligibility of the three different masker conditions was assessed in 12 normally hearing adults (ages 26–50, with six men), none of whom subsequently participated in the PET study. Conditions were presented in a randomized order. Sixteen sentences, with a total of 48 key words, were presented per condition. Sentences were presented diotically over headphones and listeners were asked to repeat back the words that they could hear from the female talker. These sentences have a very simple semantic and syntactic structure with three or four key words (e.g., the CLOWN had a FUNNY FACE). Responses were scored in terms of the number of key words correctly repeated. This was done for a range of SNRs for each masker type: 0, −3, and −6 dB SNRs for the SMN masker; −3, −6, and −9 dB SNRs for the speech masker; and −6, −9, and −12 dB SNRs for the rotated speech masker (see Fig. 2). There is a clear effect of masking condition and level on the intelligibility of the sentences. These data were used to select the SNR conditions for the PET scanning in which intelligibility was ~80%: −3 dB SNR for the SMN masker, and −6 dB SNR for the speech and rotated speech masker. Performance across the conditions at these levels was not significantly different when compared in a repeated measures analysis of variance (ANOVA) or with t-tests ($p > 0.05$). These levels were used for every presentation of the specific masking condition in the PET study.

The eight subjects for the PET study were tested prior to scanning. They were played individual BKB sentences and

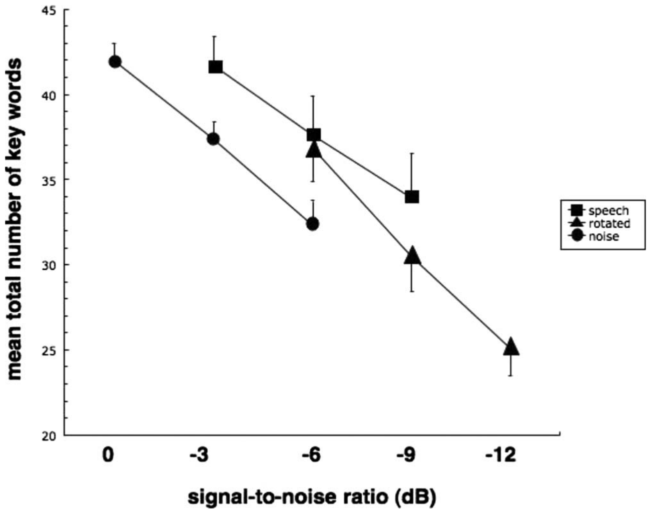

FIG. 2. Intelligibility in three different masking conditions (speech, rotated speech, and SMN), as a function of SNR, from the pilot testing conditions. The error bars show standard errors.

the masking stimuli diotically over headphones and repeated back what they could hear. Sixteen sentences, with a total of 48 key words, were presented per condition (none of which were repeated in the subsequent PET study). Intelligibility was scored by an experimenter who recorded the number of correct key words per condition as a score out of 48. This gave a score for each subject and masking condition. The order of conditions was randomized.

All of the PET subjects were able to perceive speech in the different conditions during prescan training. The average number of key words per condition was 40.4 (SD 2.61) for the speech in masking speech (=84%, with a maximum of 92% and a minimum of 75%), 39.0 (SD 2.82) for speech in masking rotated speech (=81%, with a maximum of 88% and a minimum of 73%), and 37.9 (SD 2.53) for speech-in-SMN (=79%, with a maximum of 88% and a minimum of 73%). A repeated measures ANOVA revealed that performance differed statistically across the three conditions ($F=4.62$, $df=2,7$, and $p=0.027$). *Post hoc* t-tests revealed that this arose from a significant difference between the speech and SMN masking conditions ($p=0.023$), where performance in SMN was poorer. There was no significant difference between performance on the speech-in-speech and speech-in-rotated-speech conditions, or between the speech-in-rotated-speech and speech-in-SMN conditions. The difference in intelligibility between speech-in-speech and speech-in-SMN was just over 5%.

IV. PET SCANNING

Eight right-handed native English-speaking volunteers, none of whom reported any hearing problems, were recruited and scanned. The mean age was 42, with a range 35–57. Each participant gave informed consent prior to participation in the study, which was approved by the Research Ethics Committee of Imperial College School of Medicine/Hammersmith, Queen Charlotte's & Chelsea & Acton Hospitals. Permission to administer radioisotopes was given by the Department of Health (London, UK).

PET scanning was performed with a Siemens HR++ (966) PET scanner operated in high-sensitivity three-dimensional mode. Sixteen scans were performed on each subject, using the oxygen-15-labeled water bolus technique. All subjects were scanned while lying supine in a darkened room with their eyes closed.

The stimuli were presented diotically at a comfortable level determined for each subject, and this level was kept constant over the scanning sessions. The sentence presentations began 15 s before the scanning commenced, and each sentence presented was novel (i.e., there were no repeats). As in our previous study, we used a target female talker and a male masking talker as this enabled us to give the subjects the simple instruction of "listen to the female talker." The subjects were instructed to listen passively to the female talker "for meaning" in the scanning sessions. Passive listening (i.e., with no overt responses) reduces the likelihood that activation seen is due to controlled processing aspects of the task, which would be involved if the subjects were required to make explicit responses or try and remember the sentences they heard (Scott and Wise, 2003). Such requirements have been shown to influence responses in auditory cortex (Brechmann and Scheich, 2005).

V. ANALYSIS

The images were analyzed using statistical parametric mapping (SPM99, Wellcome Department of Cognitive Neurology, http://www.fil.ion.ucl.ac.uk/spm), which allowed manipulation and statistical analysis of the grouped data. All scans from each subject were realigned to eliminate head movements between scans and normalized into a standard stereotactic space [the Montreal Neurological Institute template was used, which is constructed from anatomical magnetic resonance imaging (MRI) scans obtained on 305 normal subjects]. Images were then smoothed using an isotropic 10 mm, full width at half maximum, Gaussian kernel, to allow for variation in gyral anatomy and to improve the SNR.

VI. RESULTS

Three main contrasts were performed, both based on subtractions. In the first, regions more activated by speech-in-speech than speech-in-SMN were identified. This revealed activation confined to the left and right superior temporal gyri (STGs) (Table I), anterior to primary auditory cortex, and extending to the dorsal bank of the STS (Fig. 3). In the second, regions more activated by speech-in-rotated-speech than speech-in-SMN were identified. This revealed activation in the right STG (Table I), anterior to primary auditory cortex, and again extending to the dorsal bank of the STS (Fig. 4). Of the two peaks in this region, one lies within 2 mm in each dimension of the peak response to speech-in-speech, and thus likely represents the same peak of activation, with the spatial resolution available using PET. In the third contrast, regions more activated by the speech-in-speech than speech-in-rotated-speech were identified. This contrast did not reveal any significant activity. Finally, a conjunction analysis of both informational masking conditions

TABLE I. Peak activations for various planned contrasts.

Contrast	Region	*Z* score	*X*	*Y*	*Z*
Speech-in-speech > speech-in-SMN	Left STG	4.43	−68	−12	0
	Right STG	5.69	66	−2	−6
Speech-in-rotated-speech > speech-in-SMN	Right STG	6.53	58	−10	4
		5.34	64	−2	−4
Conjunction of speech in speech and rotated speech > speech in SMN	Right STG	6.21	64	−2	−4
		6.11	60	−10	2

revealed a peak in the right STG (Table I), which was just 2 mm more medial than the peak response in the speech-in-speech contrast, and essentially therefore reflects the same peak of activation. An additional analysis investigated any overall response to intelligibility, without regard to masker type, by using the subjects' pretesting scores as covariates across all conditions. No regions were significantly activated by this, possibly because the range of intelligibility was reduced in this study, relative to studies that expressly vary intelligibility (across all the subjects, scores ranged from 73% to 92%). In our previous study of masking, intelligibility ranged over a wider range (from 50% to 100%) and significant intelligibility related regions were seen (Scott *et al.*, 2004).

VII. DISCUSSION

Our previous study (Scott *et al.*, 2004) showed extensive bilateral superior temporal activation associated with informational masking of speech: We interpreted this as central perceptual processing of the masking speech signal, consistent with a central competition of resources in informational masking. However, we could not rule out a contribution of glimpses of the target signal in the speech masking condition as a basis of at least some of the activation, nor could we determine the nature of the central resources—acoustic or linguistic—for which there was perceptual competition. The results of the current study allow us to address these issues.

First, the activation in the speech masker condition in this study is less extensive than that seen in the previous study, suggesting that some of the changes in activation in the previous study were indeed a result of glimpses of the target signal allowed by the modulated masker. This seems to primarily affect the activations seen in more posterior audi-

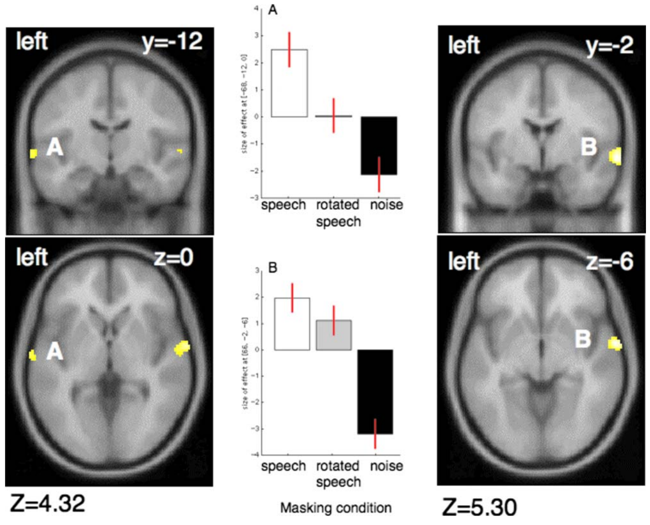

FIG. 3. (Color online) Activation for the contrast of the conditions "speech-in-speech" over the conditions "speech-in-modulated-noise" (analyzed in SPM99, $p < 0.0001$, cluster size >40 voxels). This subtraction reveals activations that are significantly greater to the masking speech than to the noise masker. The peak activations in the left and right temporal lobes are projected on the MNI TI template from SPM99: The panels on the left of the figure show the activation peak in the left hemisphere, and the panels on the right show the peak activation in the right hemisphere. The upper panels show the activation on a coronal image of the brain, and the lower panels show the activation on a transverse image. The graphs show the effect sizes as percentage signal change across conditions: While the comparison is of the activity for speech-in-speech > speech-in-noise, the activity in this peak for the speech-in-rotated-speech condition is also shown. Note that in the left temporal lobe the response to speech-in-rotated-speech is reduced relative to the response to speech-in-speech, whereas in the right temporal lobe the responses for both speech-in-speech and speech-in-rotated-speech are more similar.

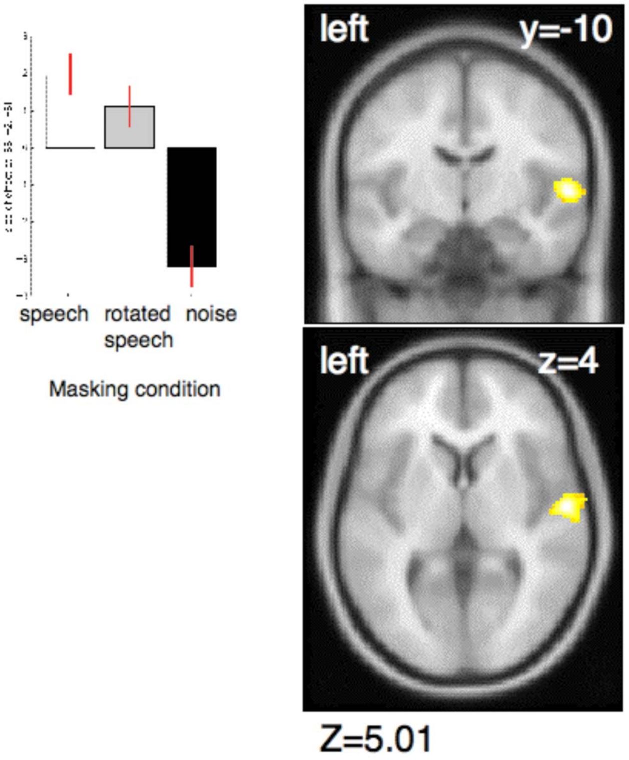

FIG. 4. (Color online) Activation for the contrast of the conditions speech-in-rotated-speech over speech-in-modulated-noise (analyzed in SPM99, $p < 0.0001$, cluster size >40 voxels). This subtraction reveals activations that are significantly greater to the masking rotated speech than to the noise masker. The peak activation in the right temporal lobe is projected on the MNI TI template from SPM99. The graph shows the effect size as percentage signal change across conditions. Note that the activation lies posterior and dorsal to the right STG peak for the speech-in-speech > speech-in-modulated-noise contrast: However, there is a subpeak (64, −2, −4, $Z = 5.34$) which lies within 2 mm of this.

tory regions: In the current study there are no peaks farther back than $y=-15$, i.e., lateral to the anterior extent of primary auditory cortex. In our previous study there were bilateral peaks extending as far back as $y=-30$, lateral to the posterior extent of primary auditory cortex. This suggests that activation associated with masking speech in anterior STG regions is not driven solely by glimpses of the target stimulus. Zatorre *et al.* (2004) recently demonstrated, by varying the number of acoustic sources perceived, that the right anterior superior temporal sulcus (STS) is associated with representation of multiple auditory objects. Potentially, it has a role in representing and selecting between competing auditory sources. Our data are consistent with this finding, suggesting a role for these anterior auditory fields in the processing of multiple auditory "streams" of information (Scott, 2005). Since these same regions are strongly implicated in the auditory processing of spoken language *without* maskers (Mummery *et al.*, 1999), these would be prime candidates for a locus of central processing of masking speech.

Second, we are able to distinguish differences between different speech masking conditions, which relate to the nature of competition for central auditory resources in masking. The presence of speech as a masker is associated with bilateral STG responses, whereas rotated speech as a masker results in right STG responses only, relative to SMN. This suggests that masking speech recruits more extensive left temporal lobe neural systems than masking rotated speech, and thus potentially different perceptual processes. Rotated speech is unintelligible, but contains much of the acoustic structure of the original speech signal, in terms of spectrotemporal dynamics, although aspects of the amplitude modulation profile are likely to be different. Importantly, since harmonics are still represented as equally spaced spectral components (although typically no longer strictly harmonic), rotated speech maintains the pitch of speech (albeit with a weaker saliency), and hence preserves the intonation of the original signal. We have previously attributed common activation of the right STG to attended (and unmasked) speech and rotated speech as a result of the presence of pitch variation in both (Scott *et al.*, 2000), and several studies explicitly investigating pitch variation have demonstrated a clear preference in the right STG/STS for such signals (Zatorre and Belin 2001; Patterson *et al.*, 2002). In contrast, functional imaging studies of intelligible speech, which compare the activation produced by speech with that produced by rotated speech, reveal solely *left* STS regions (Scott *et al.*, 2000, 2006; Narain *et al.*, 2003). It seems likely, therefore, that masking speech and rotated masking speech activate right anterior STG because of the acoustic structure and pitch variation that both have in common, while masking speech activates the left STG as a function of its intelligibility—and therefore of its linguistic status. This may have some implications for approaches to informational masking. Studies have indicated that different masking effects can be seen when a talker is masked by speech and when the masker is comprised of reversed speech—e.g., there is a release from masking when reversed speech is used as a masker (Rhebergen *et al.*, 2005), and there is a developmental profile to this, with children showing greater masking from reversed speech than adults (Johnstone and Litovsky, 2006). There can be similar effects of masking speech and masking reversed speech (Hawley *et al.*, 2004), when there are multiple maskers. One interpretation of our data is that the linguistic component of informational masking is associated with left temporal lobe mechanisms, and some of the nonlinguistic aspects of masking from speech and reversed speech may be associated with right temporal lobe mechanisms.

In conclusion, neural activity associated with masking from speech (above masking from SMN) is seen in anterior, rather than posterior, auditory fields, in areas that represent auditory objects. We also suggest that part of the masking effects of speech and rotated speech result from their shared acoustic properties. However, speech as a masker activates the left STG to a far greater degree than rotated speech (Fig. 3) because, we suggest, of its intelligibility. This is consistent with behavioral findings suggesting that informational masking effects arise from competition for both acoustic and linguistic processing resources (Brungart, 2001; Hawley *et al.*, 2004; Rhebergen *et al.*, 2005; Johnstone and Litovsky, 2006). We identify two different types of central competition in the temporal lobes, one linguistically driven in the left anterior auditory association cortex, and one driven by acoustic properties in the right anterior auditory association cortex. Such an account allows for masking by speech to show both lexical and acoustic influences, and places the neural processing of "unattended" speech in the same framework as the neural responses to attended speech (Scott *et al.*, 2000). Attention is known to modulate neural processing of sound (Brechmann and Scheich, 2005) and our results do not contradict this distinction between the two modes of response, rather they indicate that attended and unattended (masking) speech enter the processing system via the same neural pathways. Furthermore, the unattended masking speech is processed within the anteriorly directed "what" pathway of processing, which is associated with the processing of intelligible attended, unmasked speech (Scott *et al.*, 2000). Further work will determine the extent to which the right temporal lobe response to both masking speech and masking rotated speech arises from aspects of their shared acoustic structure or their pitch variation. Future studies will also be able to determine whether the left temporal response to masking speech is driven by phonological or postlexical processing of the unattended speech.

ACKNOWLEDGMENTS

S.K.S. is funded by Grant No. WT074414MA from the Wellcome Trust. Many thanks to Sam Eaton-Rosen for help in creating Fig. 1.

Blesser, B. (**1972**). "Speech perception under conditions of spectral transformation: I. Phonetic characteristics," J. Speech Hear. Res. **15**, 5–41.

Brechmann, A., and Scheich, H. (**2005**). "Hemispheric shifts of sound representation in auditory cortex with conceptual listening," Cereb. Cortex **15**, 578–587.

Brungart, D. S. (**2001**). "Informational and energetic masking effect in the perception of two simultaneous talkers," J. Acoust. Soc. Am. **109**, 1101–1109.

Byrne, D., Dillon, H., Tran, K., Arlinger, S., Wilbraham, K., Cox, R. H. B., Hetu, R. K. J., Liu, C., Kiessling, J., Kotby, M. N., Nasser, N. H. A., Elkholy, W. A. H., Nakanishi, Y. O. H., Powell, R., Stephens, D., Merdedith, R., Sirimanna, T., Tavartkiladze, G. F., Westerman, S., and Ludvigsen, C. (**1994**). "An international comparison of long-term average speech spectra," J. Acoust. Soc. Am. **96**, 2108–2120.

Festen, J., and Plomp, R. (**1990**). "Effects of fluctuating noise and interfering speech on the speech reception threshold for impaired and normal hearing," J. Acoust. Soc. Am. **88**, 1725–1736.

Foster, J. R., Summerfield, A. Q., Marshall, D. H., Palmer, L., Ball, V., and Rosen, S. (**1993**). "Lip-reading the BKB sentence lists—Corrections for list and practice effects," Br. J. Audiol. **27**, 233–246.

Hawley, M. L., Litovsky, R. Y., and Culling, J. F. (**2004**). "The benefit of binaural hearing in a cocktail party: Effect of location and type of interferer," J. Acoust. Soc. Am. **115**, 833–843.

Huckvale, M. (**2007**). "Speech filing system (software package)," http://www.phon.ucl.ac.uk/resource/sfs/ (Last viewed May, 2008).

Jacquemot, C., Pallier, C., LeBihan, D., Dehaene, S., and Dupoux, E. (**2003**). "Phonological grammar shapes the auditory cortex: A functional magnetic resonance imaging study," J. Neurosci. **23**, 9541–9546.

Johnstone, P. M., and Litovsky, R. Y. (**2006**). "Effect of masker type and age on speech intelligibility and spatial release from masking in children and adults," J. Acoust. Soc. Am. **120**, 2177–2189.

MacLeod, A., and Summerfield, Q. (**1987**). "Quantifying the contribution of vision to speech perception in noise," Br. J. Audiol. **21**, 131–141.

Mummery, C. J., Ashburner, J., Scott, S. K., and Wise, R. J. S. (**1999**). "Functional neuroimaging of speech perception in six normal and two aphasic patients," J. Acoust. Soc. Am. **106**, 449–457.

Narain, C., Scott, S. K., Wise, R. J. S., Rosen, S., Leff, A. P., Iversen, S. D., and Matthews, P. M. (**2003**). "Defining a left-lateralized response specific to intelligible speech using fMRI," Cereb. Cortex **13**, 1362–1368.

Patterson, R. D., Uppenkamp, S., Johnsrude, I. S., and Griffiths, T. D. (**2002**). "The processing of temporal pitch and melody information in auditory cortex," Neuron **36**, 767–776.

Rhebergen, K. S., Versfeld, N. J., and Dreschler, W. A. (**2005**). "Release from informational masking by time reversal of native and non-native interfering speech," J. Acoust. Soc. Am. **118**, 1274–1277.

Roucus, S., and Wilgus, A. W. (**1985**). "High quality time-scale modification for speech," in Proceedings of the IEEE International Conference of Acoustics, Speech and Signal Processing, pp. 493–496.

Scott, S. K., Rosen, S., Lang, H., and Wise, R. J. S. (**2006**). "Neural correlates of intelligibility in speech investigated with noise vocoded speech: A positron emission tomography study," J. Acoust. Soc. Am. **120**, 1075–1083.

Scott, S. K. (**2005**). "Auditory processing–speech, space and auditory objects," Curr. Opin. Neurobiol. **15**, 197–201.

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

Scott, S. K., and Johnsrude, I. S. (**2003**). "The neuroanatomical and functional organization of speech perception," Trends Neurosci. **26**, 100–107.

Scott, S. K., Rosen, S., Wickham, L., and Wise, R. J. S. (**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, S. K., and Wise, R. J. S. (**2003**). "Functional imaging and language: A critical guide to methodology and analysis," Speech Commun. **41**, 7–21.

Scott, S. K., and Wise, R. J. S. (**2004**). "The functional neuroanatomy of prelexical processing of speech," Cognition **92**, 13–45.

Wise, R. J. S., Scott, S. K., Blank, S. C., Mummery, C. J., and Warburton, E. (**2001**). "Identifying separate neural sub-systems within 'Wernicke's area,'" Brain **124**, 83–95.

Zatorre, R. J., and Belin, P. (**2001**). "Spectral and temporal processing in human auditory cortex," Cereb. Cortex **11**, 946–953.

Zatorre, R. J., Bouffard, M., and Belin, P. (**2004**). "Sensitivity to auditory object features in human temporal neocortex," J. Neurosci. **24**, 3637–3642.