Behavioral/Systems/Cognitive

# Bilateral Speech Comprehension Reflects Differential Sensitivity to Spectral and Temporal Features

**Jonas Obleser,[1] Frank Eisner,[2] and Sonja A. Kotz[1]**

[1]Max Planck Institute for Human Cognitive and Brain Sciences, 04103 Leipzig, Germany, and [2]Institute of Cognitive Neuroscience, University College London, London WC1N 3AR, United Kingdom

**Speech comprehension has been shown to be a strikingly bilateral process, but the differential contributions of the subfields of left and right auditory cortices have remained elusive. The hypothesis that left auditory areas engage predominantly in decoding fast temporal perturbations of a signal whereas the right areas are relatively more driven by changes of the frequency spectrum has not been directly tested in speech or music. This brain-imaging study independently manipulated the speech signal itself along the spectral and the temporal domain using noise-band vocoding. In a parametric design with five temporal and five spectral degradation levels in word comprehension, a functional distinction of the left and right auditory association cortices emerged: increases in the temporal detail of the signal were most effective in driving brain activation of the left anterolateral superior temporal sulcus (STS), whereas the right homolog areas exhibited stronger sensitivity to the variations in spectral detail. In accordance with behavioral measures of speech comprehension acquired in parallel, change of spectral detail exhibited a stronger coupling with the STS BOLD signal. The relative pattern of lateralization (quantified using lateralization quotients) proved reliable in a jack-knifed iterative reanalysis of the group functional magnetic resonance imaging model. This study supplies direct evidence to the often implied functional distinction of the two cerebral hemispheres in speech processing. Applying direct manipulations to the speech signal rather than to low-level surrogates, the results lend plausibility to the notion of complementary roles for the left and right superior temporal sulci in comprehending the speech signal.**

*Key words:* **speech; degraded speech; temporal processing; spectral processing; hemispheric lateralization; fMRI; noise vocoding; parametric design**

## Introduction

One hundred and fifty years after Broca's and Wernicke's endeavors and 20 years into noninvasive neuroimaging of the neural basis of language (Petersen et al., 1988), the interactive processes of the left and right cerebrum in speech comprehension remain opaque. The underlying central auditory processes are increasingly understood (Binder et al., 2000; Scott et al., 2000; Davis and Johnsrude, 2003) (for review, see Hickok and Poeppel, 2007; Zatorre and Gandour, 2008), and the often observed bilateral activations likely reflect complementary analysis steps in left and right auditory cortex, contributing differently to a unified percept of speech [for evidence from right- and left-hemispheric lesion patients, see Robin et al. (1990) and Van Lancker and Sidtis (1992)].

Zatorre generalized the idea of a left-hemispheric preference for fast temporal changes (Schwartz and Tallal, 1980) toward a trade-off in hemispheric specialization, with the right and left auditory areas being primarily suited to process spectral and temporal changes, respectively (Zatorre et al., 1992, 2002; Zatorre and Belin, 2001; Schönwiesner et al., 2005). Also, the suggestion of different temporal integration windows on which speech is processed (Poeppel, 2003) predicts a left-hemispheric preference for fast temporal changes at the cost of fine spectral resolution, and fast and slow FM modulations drive the left and right auditory cortices differentially (Boemio et al., 2005).

Generally, previous studies on auditory lateralization do not allow direct conclusions on speech. Comparably low-level acoustic stimuli and tasks, without intrinsic behavioral relevance, had to be devised to control for the trade-off between spectral and temporal characteristics as far as possible (Giraud et al., 2000, 2007; Zatorre and Belin, 2001; Hall et al., 2002; Zaehle et al., 2004; Boemio et al., 2005; Schönwiesner et al., 2005). The assumption upheld, though, was that low-level effects bear direct relevance to meaningful material, especially speech and music. Related research in nonhuman species (Rauschecker et al., 1995; Wang et al., 1995; Read et al., 2002; Fitch and Fritz, 2006; Petkov et al., 2006; Pandya et al., 2008) has limitations in generalizing to human speech processing. This gap between meaningless stimuli and a proclaimed role in the decoding of speech or music has not been closed yet (with the notable exception of intelligibility-modulating speech chimeras with varied temporal detail) (Luo and Poeppel, 2007).

It appears timely to study temporal and spectral detail in

Received March 26, 2008; revised June 24, 2008; accepted June 26, 2008.

We are grateful to Stuart Rosen (University College London, London, UK) for supplying us with the noise-vocoding code and for his help in discovering the subtleties of the algorithm. Ulrike Barth, Mandy Naumann, Annett Wiedemann, and Simone Wipper helped to acquire the behavioral and functional brain-imaging data. We are grateful to two anonymous reviewers who helped to improve this manuscript considerably with insightful questions and suggestions.

Correspondence should be addressed to Jonas Obleser, Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstrasse 1a, 04103 Leipzig, Germany. E-mail: obleser@cbs.mpg.de.

DOI:10.1523/JNEUROSCI.1290-08.2008

 0270-6474/08/288116-08$15.00/0

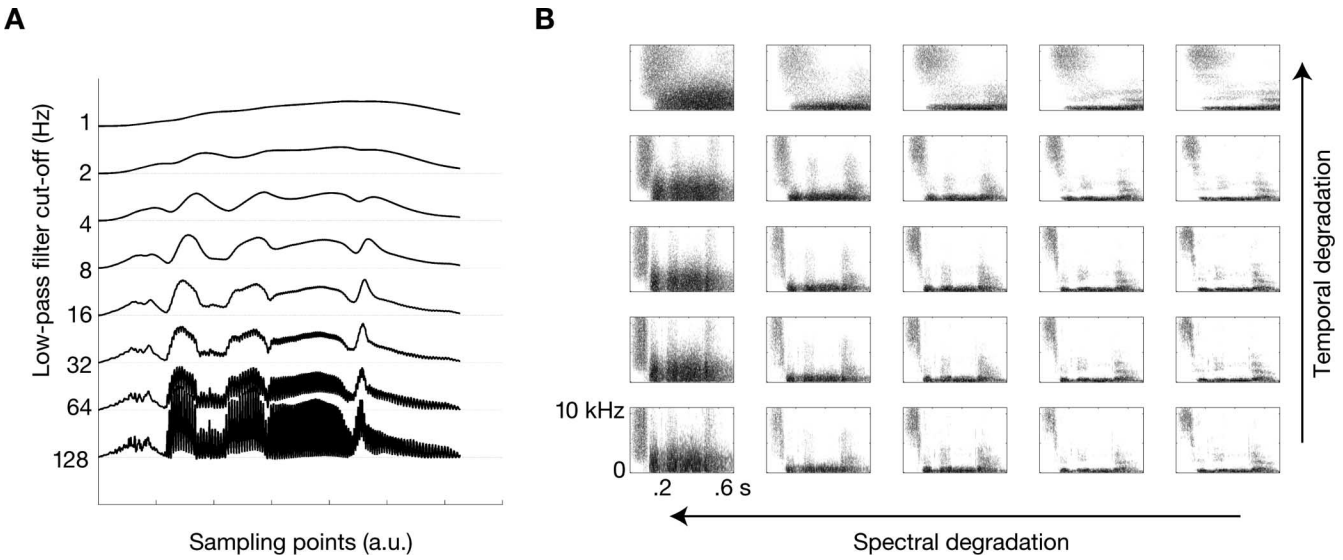


**Figure 1.** Illustration of spectral and temporal manipulations applied. ***A***, Using an exemplary speech signal, the temporal carrier of the signal as typically derived in noise-band vocoding is filtered, either retaining all natural temporal perturbations in the speech signal (e.g., using a 128 Hz low-pass cutoff, bottom waveform) or in increasing extents of low-pass filtering, which smoothes the temporal carrier and removes typical temporal detail of the speech signal. ***B***, These various filter cutoffs can be combined with the typical spectral degradation by using more or less spectral bands in which the signal is noise vocoded, yielding an orthogonal manipulation of spectral detail (number of bands for vocoding, horizontal) and temporal detail (filter cutoff by which these bands are low-pass filtered, vertical). This combination of temporal smoothing as shown in ***A*** and spectral division in varying number of bands allows for a large degree of orthogonality in degrading temporal and spectral detail, as can be observed in the various panels in ***B***. a.u., Arbitrary units.

speech comprehension directly. Using variations of noise vocoding (Shannon et al., 1995), we degraded the speech signal along two stimulus dimensions. One dimension gradually removes temporal fluctuations, whereas another removes detail of the frequency spectrum (see Materials and Methods). Applying these manipulations parametrically with multiple degradation levels of each, the listener is confronted with a highly natural task (listen to words and try to understand). Simultaneously, we tested for differential sensitivities in brain activation to changes in the temporal or the spectral detail of the speech signal. Thus, this approach can provide direct evidence for a dichotomy of pathways in the auditory analysis of speech and their supporting role in language comprehension.

## Materials and Methods

### *Participants*

Sixteen participants (8 females, age range 20–32 years) took part in the functional magnetic resonance imaging (fMRI) experiment. All were monolingual speakers of German, had normal hearing, and had no history of neurological or language-related problems. They were naive toward noise-vocoded speech and had not taken part in the pilot study. Participants were reimbursed €15. The procedure was approved of by the local ethics committee and in accordance with the declaration of Helsinki.

In the behavioral pilot study (see Fig. 2*A*), 16 different participants (8 females, age range 18–29 years) took part. They were also naive toward noise-vocoded speech and had not taken part in any previous experiments on degraded speech. They were reimbursed €7.

### *Stimulus material*

Stimuli were randomly drawn from a 560 item pool of recordings of spoken German mono-, bi-, and tri-syllabic nouns (Kotz et al., 2002). Words were recorded in a soundproof chamber by a trained female speaker and digitized at a 44.1 kHz sampling rate. Off-line editing included down-sampling to 22.05 kHz, cutting at zero-crossings before and after each word, and root mean square normalization of amplitude. The large pool of word stimuli allowed us to avoid repetition of word items both in the behavioral pilot study and in the fMRI experiment.

From each word's final audio file, various degraded versions were created using a Matlab-based noise-band-vocoding algorithm. Noise vocoding is an effective technique to manipulate the spectral detail while preserving the temporal envelope of the speech signal (Shannon et al., 1995) and render it more or less intelligible in a graded and controlled manner, depending on the number of bands used and more bands yielding a more intelligible speech signal. The technique has been used widely in behavioral and brain-imaging studies (Scott et al., 2000, 2006; Faulkner et al., 2001; Davis and Johnsrude, 2003; Obleser et al., 2007a). Usually, in noise vocoding the spectral degradation is varied by specifying the number of spectral bands to be extracted, and the extracted filter envelopes in each band are carefully smoothed (low-pass filtered) by a value that does not affect intelligibility whatsoever (e.g., 400 Hz, because the relevant temporal perturbations that contribute to intelligibility in speech seem to lie $<$20 Hz) (Xu et al., 2005).

For this study, however, we also systematically varied the temporal smoothing of the amplitude envelopes (Fig. 1*A*) over a range previously identified to have a distinct influence on intelligibility. It is noteworthy that noise vocoding allows an orthogonal manipulation of the spectral and temporal variations in any given signal that cannot be achieved with other techniques because it allows the signal to split into an arbitrary number of frequency bands and to then smooth the excitatory envelopes of each band with an arbitrary low-pass filter cutoff. For example, simply low-pass filtering a speech signal (as it has been used previously in studies of speech intelligibility) would remove fast perturbations from the envelope; but, of course, it would also affect the overall spectral content available. As becomes evident from Fig. 1*B*, this is not the case in the orthogonal manipulation approach devised here.

For the fMRI experiment, we also devised an additional baseline condition of very unintelligible stimuli (from the two-band/1 Hz condition) that were presented monaurally to the right ear. The purpose of this additional condition was twofold: first, monaural stimulation to the right ear as a strong exogenous (stimulation-dependent) lateralization factor should yield a strong contralateral effect (to the left) that can be used to put possible manipulation-dependent lateralization effects of the spectral versus temporal variations into perspective. Second, because this condition consists of highly unintelligible sounds (very low both in spectral and temporal detail), the activation it elicits serves as a functional and anatomical landmark for comprehension-independent processing, expected to activate more posterior parts of the left superior temporal

cortex [whereas the main modulations through changing intelligibility, along either the spectral or the temporal dimension, should activate mainly anterolateral superior temporal sulcus (STS) regions] (Davis and Johnsrude, 2003; Obleser et al., 2007a).

### *Design and data acquisition*

*Pilot study.* The pilot experiment consisted of 250 words, each trial randomly drawing from the pool of word stimuli (see above, i.e., each word item in each participant and each condition repetition was unique) and from the range of spectral and temporal degradation levels (2, 4, 8, 16, or 32 bands combined with 1, 2, 4, 8, or 16 Hz low-pass filtering). Participants were seated in front of a liquid crystal display screen with comfortable font size and keyboard layout. They wore headphones (HD-202; Sennheiser) and were instructed to listen to the stimuli and type in the word they had heard. The first stimulus was always a word drawn from the least degraded condition, and onscreen written feedback of the word was supplied in the first ten trials. Participants could pause at their own discretion. Scoring of responses as correct or false was based on a match of the typed string with the actual word. A mean percentage correct score was calculated for each participant and condition and submitted to a 5 × 5 repeated measures analysis with factors spectral degradation and temporal degradation. Sixteen participants completed this procedure (Fig. 2*A*), and results show highly significant main effects of spectral and temporal degradation ($F_{(3.1,46)} = 620.8, p < 0.001$; $F_{(2.7,40.8)} = 227.5, p < 0.001$). Most relevant to the intended orthogonal manipulation of intelligibility, the main effect of spectral degradation was highly significant within each level of temporal degradation (all $p < 0.001$ in separate ANOVAs; Greenhouse–Geiser corrected), as was the main effect of temporal degradation within each level of spectral degradation (two-band speech, $p < 0.045$; all other levels, $p < 0.001$; all Greenhouse–Geiser corrected). We concluded that the selection of degradation levels worked sufficiently well to be interpreted as an effective orthogonal manipulation as intended (e.g., within 16-band signals, a quasi linear increase of intelligibility with less low-pass filtering was observed; conversely, within 8 Hz filtered signals, a comparable increase of intelligibility was found across the increasing number of vocoding bands). It should be noted, though, that an interaction was also found to be significant ($F_{(6.2,93.6)} = 21.3, p < 0.001$), confirming the known different relative contribution of spectral and temporal cues (Xu et al., 2005). Given very poor spectral information (e.g., two bands), the monotonic contribution of temporal detail to comprehension, albeit significant in itself, could not elicit comprehension scores >10% correct. In contrast, comprehension at high spectral detail (e.g., 32 bands) did not fall <20% even at very poor temporal detail.

*fMRI study.* Scanning was performed using a Siemens 3T scanner (Siemens) with birdcage headcoil. Participants were comfortably positioned in the bore and wore air-conduction headphones (Resonance Technology). After a brief (15 trial) familiarization period, the actual experiment was started.

Participants were required to listen attentively to the stimuli words and to indicate by way of a four-way button system how comprehensible this trial's word had been. This rating technique shows remarkable consistency with actual recognition scores [see also the study by Davis and Johnsrude (2003) in which rating and recognition scores within participants showed a correlation score of 0.98] and was used in-scanner for its simplicity and efficiency.

Functional scans were acquired every 9 s, with a stimulus being presented 5.5 s before each scan (sparse temporal sampling) (Edmister et al., 1999; Hall et al., 1999). Two seconds before each scan (i.e., 3.5 s after word onset), a question mark appeared on the screen, prompting participants to press a button to rate the intelligibility of the heard stimulus.

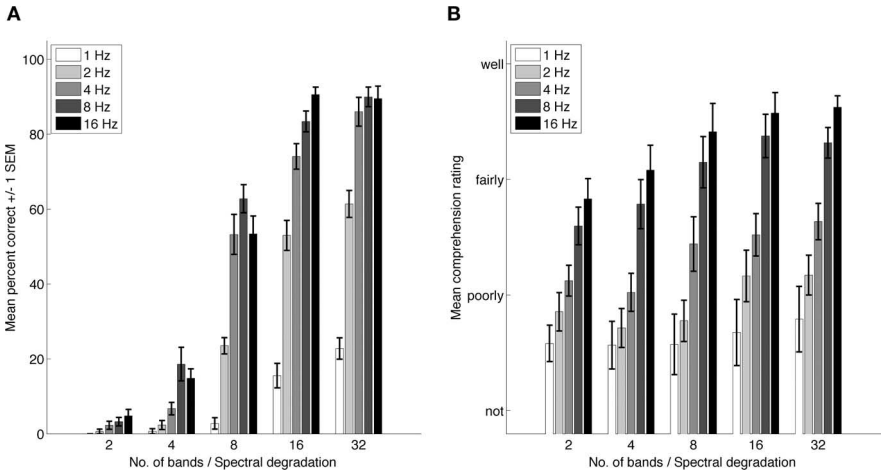


**Figure 2.** ***A***, ***B***, Results of behavioral pilot sample (***A***) and in-scanner testing (***B***). In both graphs, gray shading going from light to dark indicates increasing (more natural) temporal detail, whereas bar groups indicate increasing (more natural) spectral detail. In behavioral testing (left), a comprehension (word typing) task was used, whereas for in-scanner testing, a four-way rating of comprehensibility was applied (right). See Materials and Methods and Results for statistical details.

Digit/hand-to-button assignment was balanced across the participant sample. For the rarely occurring silent (off) trials, participants were instructed to press any button of their choice after the prompt. Figure 2*B* shows that the in-scanner task was accomplished well and yielded highly concordant results to the pilot word recognition study performed by a different sample.

Each trial's stimulus was drawn pseudorandomly from 1 of the 25 conditions (5 spectral degradation levels × 5 temporal degradation levels). It should be noted that we conceived of this study as a parametric variation study, and we therefore planned with a comparably large number of levels (5 levels) in each of the two parameters. The 25 experimental cells of spectral and temporal degradation had 9 trials each (amounting to 225 trials). With an additional 18 trials of the monaural unintelligible baseline condition (see above, Stimulus material) and 18 trials of silence (no stimulus), a total number of 261 trials and MR volume scans were recorded, lasting ~35 min.

Echoplanar imaging scans were acquired in 26 axial slices covering the entire brain with an in-plane resolution of $3 \times 3\ mm^2$ and a 3 mm slice thickness (repetition time = 9 s; acquisition time = 2000 s; excitation time = 30 ms; flip angle 90°; field of view 192 mm; matrix size 64; interleaves slice acquisition and no gap between slices). For each participant, the individual high-resolution three-dimensional T1-weighted MR scan acquired in a previous session was available for normalization, coregistration, and data visualization.

### *Data analysis*

Functional data were motion-corrected off-line with the Siemens motion correction protocol (Siemens). Further analyses were performed using Statistical Parametric Mapping 5 (SPM5; Wellcome Imaging Department, University College London, London, UK). fMRI time series were resampled to a $2\ mm^3$ voxel size, corrected for field inhomogeneities ("unwarped"), normalized (by segmenting each participant's T1-weighted image according to the SPM5 gray-matter template image and using the parameters gained for normalizing this participant's fMRI images) (Warren et al., 2006; Obleser et al., 2007a), and smoothed using an isotropic $8\ mm^3$ kernel.

In each participant, a general linear model using two regressors of interest (spectral variation, temporal variation) based on the parametric variation of these two stimulus dimensions (both coded as ranging from 1 to 5) was estimated with a finite impulse response basis function (order 1, window length 1). Estimates of both conditions from all 16 participants were then submitted to second-level one-tailed *t* tests.

All reported group SPM statistics were thresholded at $p < 0.05$ (corrected for familywise errors, that is, multiple comparisons based on the number of voxels), if not indicated otherwise. To avoid the introduction

of spurious multicollinearity between the spectral and temporal regressors (which were perfectly noncorrelated) into our model (cf. Petersson et al., 1999), the occurrence of monaural trials had to be modeled in separate first- and second-level models after the same principals and thresholds.

To quantify any extent of lateralization difference between spectrally and temporally driven brain activations, calculation of a lateralization quotient (LQ) was planned. After applying the (very strict) statistical threshold described above, we compared the peak activation strength ($Z$) of left- and right-hemispheric activation clusters weighted by their cluster extent ($k$) (in number of voxels) by calculating the LQ as $(Z_{left} \times k_{left} - Z_{right} \times k_{right})/(Z_{left} \times k_{left} + Z_{right} \times k_{right})$.

To obtain a measure of across-sample reliability of the lateralization, we also applied a jack-knife procedure (Efron and Tibshirani, 1993) in which the SPM random-effects tests for spectral and temporal variation were rerun $n$ times (with $n$ being the number of participants, 16) while omitting one participant at a time. Hence, $n$ measures of LQ were calculated. The procedure allows estimating the reliability and variability of any statistical effect (for applications of jack-knifing, see Biswal et al., 2001; Ulrich and Miller, 2001; Obleser et al., 2006). Yielding $n$ models with $n - 1$ participants, it preserved a reasonably good signal-to-noise ratio and all advantages of second-level modeling because it consisted of only one participant less and was therefore preferable over estimates of lateralization in single participants' models with their vastly reduced signal-to-noise ratio.

For further region of interest analyses, the MaRsbar (MARSeille Boîte À Région d'Intérêt) toolbox (Brett et al., 2002) was used to extract individual and condition-specific values of local percentage signal change in the activation clusters as defined on the basis of the aforementioned SPM5 group statistics results. These values were then submitted to conventional repeated measures ANOVAs and *post hoc* paired $t$ tests.

## Results

### Behavioral data

As the pilot study had indicated (see Materials and Methods), the in-scanner comprehension rating yielded very strong main effects of both the spectral and the temporal intelligibility manipulation ($F_{(1.1,16.8)} = 46.2$, $p < 0.001$; $F_{(1.5,22.7)} = 25.7$, $p < 0.001$) (Fig. 2). Also, highly comparable with the pilot results from a different sample, all one-way ANOVAs for main effects of either spectral or temporal manipulation within a given degradation level (i.e., of the correspondingly other manipulation domain) attained significance (within two-band speech, $p < 0.01$; all other ANOVAs, $p < 0.001$).

### fMRI results

All participants showed extensive bilateral activation of the temporal cortices in response to sound and were included in the ensuing group analyses.

At the threshold determined a priori, both the spectral and the temporal parametric variations appeared to vigorously drive brain bilateral areas in the anterior superior temporal cortex. Whereas both parametric variations elicited bilateral peak activations in the mid to anterolateral STS (Fig. 3, Table 1), the relative pattern of activations showed a clear-cut dissociation of hemispheric balance and spectral versus temporal variation (Figs. 3, 4). The spectral variation, although overall being the slightly stronger influence on anterolateral temporal activation with broader and stronger clusters, exhibited a stronger right- than left-hemispheric peak activation (right, $Z = 6.09$; left, $Z = 5.98$), whereas the temporal variation exhibited the opposite pattern, that is, a stronger left-hemispheric peak (right, $Z = 5.41$; left, $Z = 5.61$). In both variations, stronger peak activations were accompanied by larger cluster sizes (Table 1). Also, the peak voxels of spectral and temporal sensitivity were 10.2 mm apart in the right hemisphere (the spectral peak being located more posterior and superior) (compare Table 1, Fig. 4*A*), which amounts to more than three times the acquired voxel size; in the left hemisphere, this difference was not as pronounced (5.6 mm).

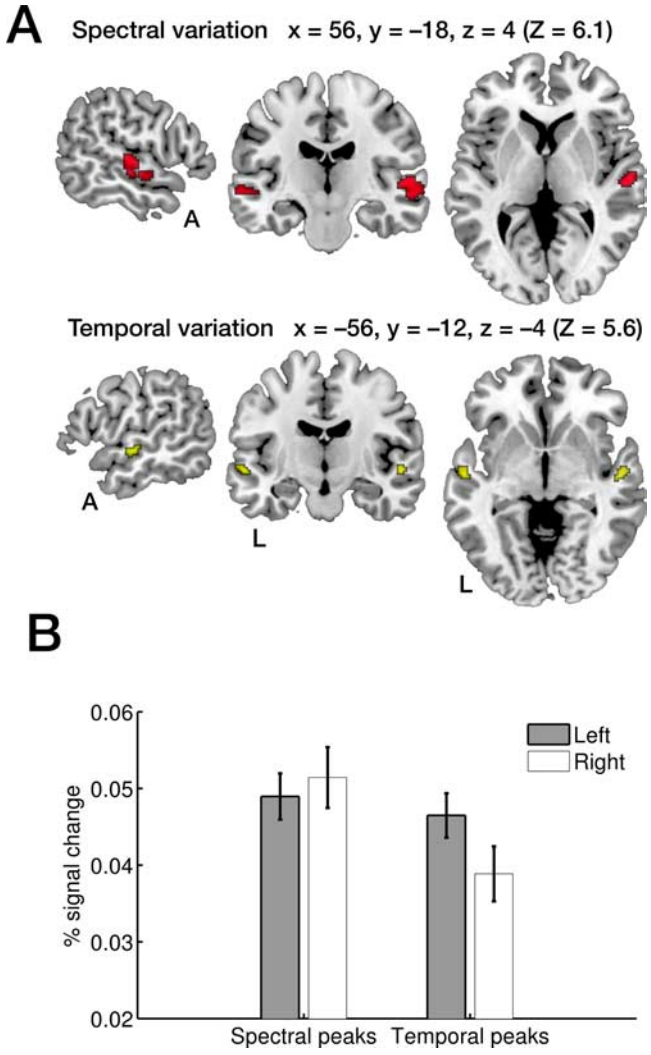


**Figure 3.** Display of spectral and temporal variation sensitivity. ***A***, Group statistical results from parametric SPM analyses (see Results for details) thresholded at $p < 0.05$ (familywise-error correction) are shown for the spectral sensitivity regressor (top) and the temporal sensitivity regressor (bottom). For both conditions, the overlays on sagittal, coronal, and axial (from left to right) slices of a T1-weighted brain template are shown for the respective peak coordinate. L, Left; A, anterior. ***B***, Bar graph reflecting the significant interaction of spectral-temporal sensitivity and left-right hemispheric preference, as observed in the regions of interest analysis.

### Region of interest analyses

A series of interesting effects was found when analyzing single participants' percentage signal change values of the spectral and temporal variation regressors extracted from the random-effects peak sites of spectral and temporal variation in both hemispheres.

Most importantly, the interaction of hemisphere and spectral/temporal peak activations was confirmed (peak × hemisphere, $F_{(1,15)} = 11.0$, $p = 0.005$) (Fig. 3*B*). Percentage signal change differences between the spectral and the temporal peak site (as derived from the spectral and the temporal random-effects test) (Table 1) were very pronounced in the right hemisphere, where the spectral peak yielded stronger signal change values than the more anterior and inferior temporal peak ($t_{(15)} = 4.25$, $p <$

0.0001), whereas no difference in the left hemisphere was observed (Fig. 3*B*).

Also, as expected from the random-effects *Z* values, spectral variation in the stimulus material generally appeared to drive the activation in the superior temporal sulci more effectively: the spectral regressor yielded stronger signal change than the temporal regressor ($F_{(1,15)} = 13.3$, $p < 0.005$), and signal change values in both spectral peaks regardless of hemisphere were stronger than those extracted from the temporal peaks ($F_{(1,15)} = 11.6$, $p < 0.005$).

**Table 1. Overview of significant clusters in random-effects analysis (familywise-error corrected, $p < 0.05$)**

| Site | MNI coordinates | | | *Z* | Extent (mm³) |
|---|---|---|---|---|---|
| Spectral variation | | | | | |
| Right middle lateral STG (BA 22) | 56 | −18 | 4 | 6.09 | 2432 |
| Left middle STG/STS (BA 22) | −60 | −8 | −4 | 5.98 | 1664 |
| Temporal variation | | | | | |
| Left middle STS (BA 21) | −56 | −12 | −4 | 5.61 | 312 |
| Right middle STG/STS (BA 22) | 54 | −12 | −4 | 5.41 | 232 |
| Monaural stimulation (right ear, control) | | | | | |
| Left posterior STG (BA 42) | −58 | −34 | 10 | 5.5 | 376 |
| Correlation with comprehension rating | | | | | |
| Right transverse temporal gyrus (BA 42) | 62 | −12 | 8 | 4.99 | 3632 |
| Left MTG | −50 | −36 | −6 | 4.75 | 776 |
| Left IFG (BA 9) | −48 | 4 | −22 | 4.57 | 520 |
| Left STG (BA 22) | −62 | −16 | 0 | 4.57 | 1688 |

Specifications refer to peak voxels. IFG, Inferior frontal gyrus; MTG, middle temporal gyrus.

### Reliability and extent of lateralization

The LQ for the spectral and the temporal variation, which take into account both the peak *Z* value and the cluster extent from the random-effects models (see Materials and Methods) and control for overall differences in activation strength, also reflected the hemispheric imbalance. Spectral variation in the speech signal revealed a mild lateralization to the right, LQ = −0.20. In contrast, we observed an LQ of +0.17 (i.e., lateralization to the left) for the temporal variation.

The subsequent jack-knife procedure (see Materials and Methods) was used to estimate the reliability of these lateralization estimates. It confirmed a stable pattern of results indicated by small error bars (Fig. 4*B*). Specifically, the observed lateralization to the left for temporal signal fluctuations appeared less stable (i.e., more dependent on the subset of participants under consideration in the jack-knife procedure), whereas the mild lateralization to the right for spectral intelligibility variations in the signal proved to be very reliable.

Analysis of the right-ear monaural stimulation as an estimate for the extent of exogenous (stimulation-dependent) lateralization confirmed the sensitivity of our set up to detect differences in hemispheric lateralization. By the same rigorous threshold, it yielded strong lateralization to the left, with one peak in posterior superior temporal gyrus (STG). This monaural stimulation, which also consisted of highly unintelligible two-band, 1 Hz low-pass filtered words, also clearly peaked more posterior and superior than the STS activation observed for spectral and temporal change sensitivity ($Z = 5.5$; MNI coordinates, −58, −34, 10; most likely located in the left planum temporale) (Westbury et al., 1999) (Fig. 4*A*, Table 1).

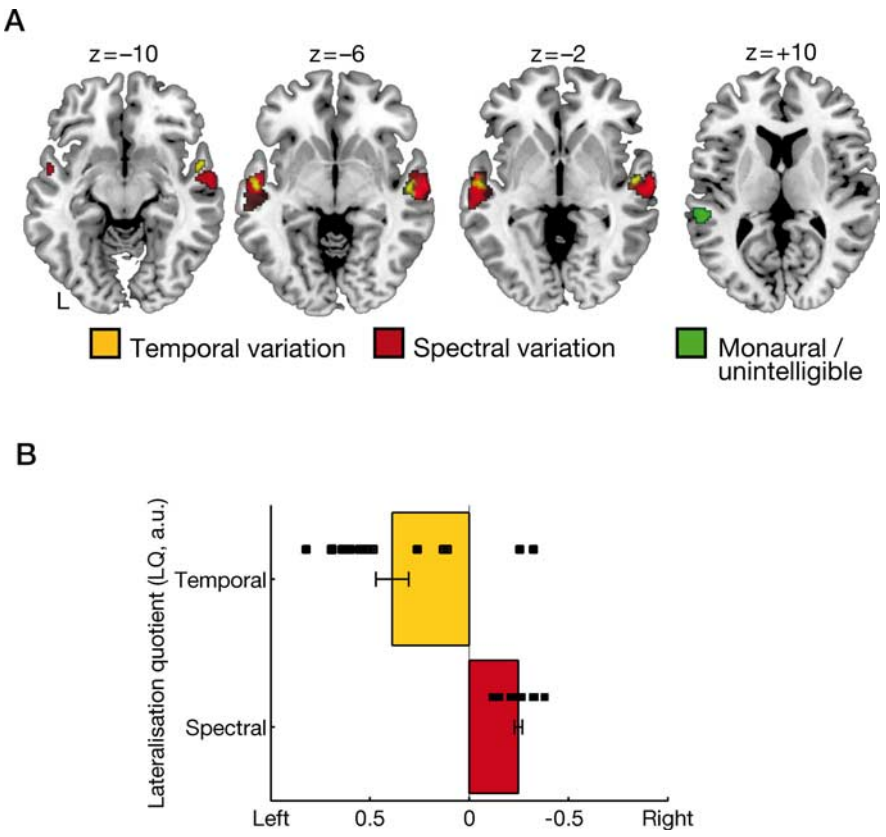


**Figure 4.** Lateralization and functional hierarchy of spectral, temporal, and monaural conditions. ***A***, Four axial slices through a T1-weighted brain template (left to right slices going from inferior to superior) show overlays of spectral (red) and temporal (yellow) sensitivity as well as the monaural unintelligible condition (green), all thresholded at $p < 0.05$ (familywise-error correction). L, Left. ***B***, Result of the iterative (jack-knife) reanalysis of the group fMRI data's lateralization quotient (see Materials and Methods for detail). Single data points indicate results of iterations, with one subject left out at a time (hence, $n = 15$ for all analyses). Note that the mild lateralization to the right (for spectral sensitivity; red) is very stable across the sample, whereas the left lateralization (for temporal sensitivity; yellow) exhibits greater variability depending on the sample configuration. a.u., Arbitrary units.

### Correlation with comprehension

Last, we used the behavioral data gathered during scanning, which indicated participants' rating of how comprehensible a degraded word had been, as a predictor in modeling the fMRI data. We modeled each participant's ratings and brain data and submitted the resulting contrasts to a random-effects test. The results showed that the rated comprehension predicted the activation strength in bilateral superior temporal cortex almost as effectively ($Z < 5$) as the actual degradation levels, although the activation peak was observed more superiorly, that is, in Heschl's gyrus [Brodmann's area 42 (BA 42)] rather than in STS. Also,

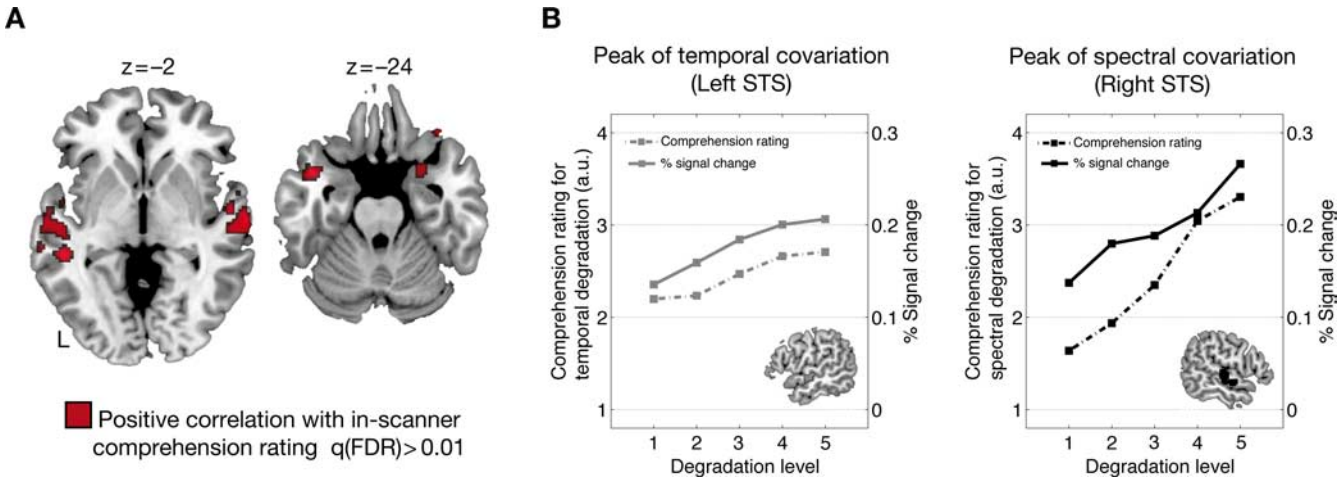


**Figure 5.** Relationship of comprehension ratings and brain activation. ***A***, Two axial slices approximately through the STS (left) and the middle/inferior temporal gyrus (right) display the maxima of positive correlation with comprehension scores (see Results for details), thresholded at $q(\mathrm{FDR}) > 0.01$. FDR, False discovery rate; L, left. ***B***, For the peak regions of interest for maximal temporal (left) and maximal spectral (right) sensitivity, the monotonic increase in both the percentage signal change estimates (solid lines) and the corresponding comprehension rating scores (dashed lines) for each of the five (temporal and spectral, respectively) degradation levels is plotted. Note that the steeper response curve in the spectral conditions corroborates the steeper change in comprehension rating scores. a.u., Arbitrary units.

rating-related activation extended deeply into the medial and anterior sections of the temporal lobe (Fig. 5); stronger activation was observed whenever comprehension was likely to succeed.

## Discussion

Is the bihemispheric processing rooted in different sensitivities to temporal and spectral changes in the speech signal? This investigation was devised to answer this question, using independent manipulations of the temporal and the spectral detail of spoken words in a listening task and testing for brain regions most sensitive to the trial-to-trial fluctuations along either the temporal or the spectral dimension.

The main results of our study are the minute differences in sensitivity of the left and right auditory areas to temporal and spectral fluctuations and their direct link to speech comprehension.

This is in line with suggestions by Zatorre et al. (2002) and Poeppel (2003). It is particularly in agreement with the concept of spectral fine-tuning of the right auditory cortex and its importance in tracking and processing pitch contours in speech (Kotz et al., 2003; Gandour et al., 2004) or the specific role for pitch in tone languages (for review, see Zatorre and Gandour, 2008).

Both temporal and spectral changes were effective in driving the hemodynamics of the mid to anterolateral superior temporal sulci. In both hemispheres, a monotonic increase of the BOLD response was observed for increasingly intelligible stimuli, that is, for more of either the natural temporal or the spectral detail available in the acoustic signal. Pertaining to lateralization, three points deserve further elaboration as follows.

First, the observed subtle shifts in hemispheric preference when using a temporal-change predictor (to the left) or a spectral-change predictor (to the right) point in the expected direction from theoretical considerations as well as previous, more low-level (i.e., nonspeech) studies (Giraud et al., 2000; Zatorre and Belin, 2001; Boemio et al., 2005; Schönwiesner et al., 2005; see also Giraud et al., 2007). The result of mild lateralization is also substantiated by the jack-knife reanalysis, which one allows to estimate the across-sample reliability in this moderate lateralization effect.

Second, our data show a stable coupling of spectral change with right STS activity. Scott et al. (2000) had noted previously this sensitivity to spectral changes in the right STS. It was found to be equally sensitive to (unintelligible) spectrally rotated and unmanipulated speech (both of which had preserved natural pitch variation) and to manipulations of the spectral envelope information (Warren et al., 2005). Here, the right-hemispheric STS differentiated robustly between spectral and temporal manipulations. Also, the two different stimulus dimensions showed a topographical shift between their right-hemispheric peak locations of several voxel sizes (Fig. 4). Right auditory areas might execute a specialized supportive function, namely a tuning to signal changes on a somewhat slower time scale and requiring longer time windows of integration (cf. Luo and Poeppel, 2007). The relatively higher consistency of the right-hemisphere responses (Fig. 4*B*) speaks to a stronger specialization: less variability in responses most likely indicates less variability in the neuronal populations and their receptive field properties, also implying a more constrained computational role of right-hemispheric auditory areas in speech comprehension. Thus, our data ultimately might be interpreted as a case for a right-hemisphere specialization.

In contrast, spectral and temporal changes were equally effective in driving the left STS. One might, therefore, argue that, in contrast to right auditory areas with their tight coupling to aspects of spectral processing, left auditory areas play a more versatile role in speech signal processing, being equally sensitive to all intelligibility-modulating changes in the speech signal. Whether such a broader tuning would justify claiming supremacy for the left in speech signal processing, however, is beyond the scope of this experiment.

Third, our results underline the integrative manner by which both auditory cortices operate on the speech signal. The overall bilateral activation is concordant with an entire series of studies on speech comprehension (i.e., most of the functional neuroimaging studies cited above report bilateral STS activations) and is warranted by current theoretical reflections on the functional neuroanatomy of speech comprehension (Friederici and Alter, 2004; Hickok and Poeppel, 2007; Zatorre and Gandour, 2008). Essentially, these models agree that the right hemisphere is in-

strumental in analyzing paralinguistic information of speech, such as emotional prosody or talker identity. These aspects, in turn, are conveyed on suprasegmental time scales and are tied closely to broad spectral rather than fine temporal detail (as in our manipulations).

Generally, all our main observations (mild lateralization effects, predominance of spectral sensitivity, bilaterality) are in line with the endogenous differences in oscillatory rhythms in left and right auditory cortices reported recently (Giraud et al., 2007). This combined EEG and fMRI study explored default (i.e., stimulation-independent) auditory networks and demonstrated a mild difference in "tuning functions" of left and right Heschl's gyrus. The right hemisphere leans strongly toward slow oscillations in the theta (3–6 Hz) range, whereas the left (not as strongly) leans toward gamma-range oscillations (28–40 Hz), again in line with the proposed theoretical frameworks (Poeppel, 2003; Zatorre and Gandour, 2008).

Experimental designs such as this, in which two signal dimensions are manipulated independently, show that strong bilateral activation through intelligible and meaningful signals (as seen here and in numerous previous neuroimaging studies) does not necessarily indicate that identical processes are operating in the left and the right. Both auditory cortices are strongly activated by speech signals, but their activations can reflect the processing of different signal dimensions, an observation only possible when different levels of such dimensions are tested in a parametric and independent manner.

Our data also speak to functional hierarchies of auditory processing. Data from speech comprehension commonly yield peak activations hierarchically downstream to Heschl's gyrus (Rauschecker and Tian, 2000; Kumar et al., 2007; Obleser et al., 2007b). We see activation in the STS rather than in the superior temporal plane, unlike studies with less meaningful auditory stimuli (Zatorre and Belin, 2001; Schönwiesner et al., 2005) (compare Fig. 3).

After anatomical probability mappings, the activations for spectral and temporal detail lie inferior and anterior of primary auditory cortex (Rademacher et al., 2001). Not only did the right-ear monaural condition show strong lateralization to the left midposterior STG, but the observed peak for these unintelligible signals also originated substantially more posterior (planum temporale). This also corroborates a stream of activation for increasingly comprehensible and meaningful speech signals from more primary areas in the supratemporal plane into more lateral, anterior, and inferior areas (anterior STG, STS) (Binder et al., 2000; Scott et al., 2000; Davis and Johnsrude, 2003; Obleser et al., 2007a).

As for a related hierarchy of form (in)dependence, note that our peak activations for (form-dependent) sensitivity to spectral or temporal detail are, on average, not as inferior as the main activations reported in studies claiming form independence of their intelligible variations (Scott et al., 2000) [Davis and Johnsrude (2003) report their activations in middle temporal gyrus].

Last, the behavioral responses from pretests (in a different sample) as well as in the scanner are important in understanding how spectral and temporal changes impact the perception of the speech signal. Although comprehension, expectedly, reaches ceiling with 16-band spectral detail (for review, see Shannon et al., 2004), interesting differences in the gain functions from temporal and spectral detail occurred (compare Fig. 2), which are in line with previous behavioral studies (Fu and Shannon, 1999; Xu et al., 2005) and point to a predominance of spectral cues.

We can test whether these dimensions affect word comprehensibility and brain activation states alike. The closest coupling of BOLD and comprehensibility changes appeared deep in the medial and anterior sections of the temporal lobe (Fig. 5*A*), regions supposedly involved in semantic retrieval processing (Devlin et al., 2002; Halgren et al., 2006). Another striking parallel was the increased efficacy of frequency spectrum: not only were spectral changes more powerful in driving the measured BOLD changes, they also yielded steeper changes in participants' comprehension ratings. Although correlation measures cannot quantify this observation (both temporal and spectral changes as well as both ratings were strictly monotonic functions, i.e., all rank correlation coefficients amounted to 1), Figure 5*B* gives an impression of these parallels: the steeper change in the spectral-related BOLD variation corresponds to the comparably steeper comprehensibility ratings for this manipulation. In keeping with previous behavioral studies (Fu and Shannon, 1999; Xu et al., 2005) and our own pilot study results, the temporal manipulations yield a flatter rating curve, which in turn is paralleled by less modulation of the BOLD signal change.

Using variations of natural speech, we have demonstrated a subtle hemispheric asymmetry in the sensitivity to spectral and temporal detail in speech. Degrading spectral information in the speech signal affected the hemodynamic response in the right STS more than a loss of temporal information, whereas a loss of temporal detail most strongly affected the response of the left STS. This information selectivity was overall more pronounced in the right hemisphere. Our results directly support an account of speech processing in which both auditory cortices serve complimentary functions, based on differently weighted temporal integration windows or spectrotemporal feature sensitivities, and speak against a simple left hemisphere bias in speech and language comprehension.

## References

Binder JR, Frost JA, Hammeke TA, Bellgowan PS, Springer JA, Kaufman JN, Possing ET (2000) Human temporal lobe activation by speech and nonspeech sounds. Cereb Cortex 10:512–528.

Biswal BB, Taylor PA, Ulmer JL (2001) Use of jackknife resampling techniques to estimate the confidence intervals of fMRI parameters. J Comput Assist Tomogr 25:113–120.

Boemio A, Fromm S, Braun A, Poeppel D (2005) Hierarchical and asymmetric temporal sensitivity in human auditory cortices. Nat Neurosci 8:389–395.

Brett M, Anton JL, Valabregue R, Poline JP (2002) Region of interest analysis using an SPM toolbox [abstract]. Paper presented at the 8th International Conference on Functional Mapping of the Human Brain, Sendai, Japan, June.

Davis MH, Johnsrude IS (2003) Hierarchical processing in spoken language comprehension. J Neurosci 23:3423–3431.

Devlin JT, Russell RP, Davis MH, Price CJ, Moss HE, Fadili MJ, Tyler LK (2002) Is there an anatomical basis for category-specificity? Semantic memory studies in PET and fMRI. Neuropsychologia 40:54–75.

Edmister WB, Talavage TM, Ledden PJ, Weisskoff RM (1999) Improved auditory cortex imaging using clustered volume acquisitions. Hum Brain Mapp 7:89–97.

Efron B, Tibshirani RJ (1993) An introduction to the bootstrap. New York: Chapman and Hall/CRC.

Faulkner A, Rosen S, Wilkinson L (2001) Effects of the number of channels and speech-to-noise ratio on rate of connected discourse tracking through a simulated cochlear implant speech processor. Ear Hear 22:431–438.

Fitch WT, Fritz JB (2006) Rhesus macaques spontaneously perceive formants in conspecific vocalizations. J Acoust Soc Am 120:2132–2141.

Friederici AD, Alter K (2004) Lateralization of auditory language functions: a dynamic dual pathway model. Brain Lang 89:267–276.

Fu QJ, Shannon RV (1999) Recognition of spectrally degraded and

frequency-shifted vowels in acoustic and electric hearing. J Acoust Soc Am 105:1889–1900.
Gandour J, Tong Y, Wong D, Talavage T, Dzemidzic M, Xu Y, Li X, Lowe M (2004) Hemispheric roles in the perception of speech prosody. Neuroimage 23:344–357.
Giraud AL, Lorenzi C, Ashburner J, Wable J, Johnsrude I, Frackowiak R, Kleinschmidt A (2000) Representation of the temporal envelope of sounds in the human brain. J Neurophysiol 84:1588–1598.
Giraud AL, Kleinschmidt A, Poeppel D, Lund TE, Frackowiak RS, Laufs H (2007) Endogenous cortical rhythms determine cerebral specialization for speech perception and production. Neuron 56:1127–1134.
Halgren E, Wang C, Schomer DL, Knake S, Marinkovic K, Wu J, Ulbert I (2006) Processing stages underlying word recognition in the anteroventral temporal lobe. Neuroimage 30:1401–1413.
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.
Hall DA, Johnsrude IS, Haggard MP, Palmer AR, Akeroyd MA, Summerfield AQ (2002) Spectral and temporal processing in human auditory cortex. Cereb Cortex 12:140–149.
Hickok G, Poeppel D (2007) The cortical organization of speech processing. Nat Rev Neurosci 8:393–402.
Kotz SA, Cappa SF, von Cramon DY, Friederici AD (2002) Modulation of the lexical-semantic network by auditory semantic priming: an event-related functional MRI study. Neuroimage 17:1761–1772.
Kotz SA, Meyer M, Alter K, Besson M, von Cramon DY, Friederici AD (2003) On the lateralization of emotional prosody: an event-related functional MR investigation. Brain Lang 86:366–376.
Kumar S, Stephan KE, Warren JD, Friston KJ, Griffiths TD (2007) Hierarchical processing of auditory objects in humans. PLoS Comput Biol 3:e100.
Luo H, Poeppel D (2007) Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. Neuron 54:1001–1010.
Obleser J, Scott SK, Eulitz C (2006) Now you hear it, now you don't: transient traces of consonants and their nonspeech analogues in the human brain. Cereb Cortex 16:1069–1076.
Obleser J, Wise RJ, Alex Dresner M, Scott SK (2007a) Functional integration across brain regions improves speech perception under adverse listening conditions. J Neurosci 27:2283–2289.
Obleser J, Zimmermann J, Van Meter J, Rauschecker JP (2007b) Multiple stages of auditory speech perception reflected in event-related fMRI. Cerebral Cortex 17:2251–2257.
Pandya PK, Rathbun DL, Moucha R, Engineer ND, Kilgard MP (2008) Spectral and temporal processing in rat posterior auditory cortex. Cereb Cortex 18:301–314.
Petersen SE, Fox PT, Posner MI, Mintun M, Raichle ME (1988) Positron emission tomographic studies of the cortical anatomy of single-word processing. Nature 331:585–589.
Petersson KM, Nichols TE, Poline JB, Holmes AP (1999) Statistical limitations in functional neuroimaging. II. Signal detection and statistical inference. Philos Trans R Soc Lond B Biol Sci 354:1261–1281.
Petkov CI, Kayser C, Augath M, Logothetis NK (2006) Functional imaging reveals numerous fields in the monkey auditory cortex. PLoS Biol 4:e215.
Poeppel D (2003) The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time'. Speech Commun 41:245–255.
Rademacher J, Morosan P, Schormann T, Schleicher A, Werner C, Freund HJ, Zilles K (2001) Probabilistic mapping and volume measurement of human primary auditory cortex. Neuroimage 13:669–683.
Rauschecker JP, Tian B (2000) Mechanisms and streams for processing of "what" and "where" in auditory cortex. Proc Natl Acad Sci U S A 97:11800–11806.
Rauschecker JP, Tian B, Hauser M (1995) Processing of complex sounds in the macaque nonprimary auditory cortex. Science 268:111–114.
Read HL, Winer JA, Schreiner CE (2002) Functional architecture of auditory cortex. Curr Opin Neurobiol 12:433–440.
Robin DA, Tranel D, Damasio H (1990) Auditory perception of temporal and spectral events in patients with focal left and right cerebral lesions. Brain Lang 39:539–555.
Schönwiesner M, Rübsamen R, von Cramon DY (2005) Hemispheric asymmetry for spectral and temporal processing in the human antero-lateral auditory belt cortex. Eur J Neurosci 22:1521–1528.
Schwartz J, Tallal P (1980) Rate of acoustic change may underlie hemispheric specialization for speech perception. Science 207:1380–1381.
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, 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.
Shannon RV, Fu QJ, Galvin J 3rd (2004) The number of spectral channels required for speech recognition depends on the difficulty of the listening situation. Acta Otolaryngol Suppl:50–54.
Ulrich R, Miller J (2001) Using the jackknife-based scoring method for measuring LRP onset effects in factorial designs. Psychophysiology 38:816–827.
Van Lancker D, Sidtis JJ (1992) The identification of affective-prosodic stimuli by left- and right-hemisphere-damaged subjects: all errors are not created equal. J Speech Hear Res 35:963–970.
Wang X, Merzenich MM, Beitel R, Schreiner CE (1995) Representation of a species-specific vocalization in the primary auditory cortex of the common marmoset: temporal and spectral characteristics. J Neurophysiol 74:2685–2706.
Warren JD, Jennings AR, Griffiths TD (2005) Analysis of the spectral envelope of sounds by the human brain. Neuroimage 24:1052–1057.
Warren JE, Sauter DA, Eisner F, Wiland J, Dresner MA, Wise RJ, Rosen S, Scott SK (2006) Positive emotions preferentially engage an auditory-motor "mirror" system. J Neurosci 26:13067–13075.
Westbury CF, Zatorre RJ, Evans AC (1999) Quantifying variability in the planum temporale: a probability map. Cereb Cortex 9:392–405.
Xu L, Thompson CS, Pfingst BE (2005) Relative contributions of spectral and temporal cues for phoneme recognition. J Acoust Soc Am 117:3255–3267.
Zaehle T, Wüstenberg T, Meyer M, Jäncke L (2004) Evidence for rapid auditory perception as the foundation of speech processing: a sparse temporal sampling fMRI study. Eur J Neurosci 20:2447–2456.
Zatorre RJ, Belin P (2001) Spectral and temporal processing in human auditory cortex. Cereb Cortex 11:946–953.
Zatorre RJ, Gandour JT (2008) Neural specializations for speech and pitch: moving beyond the dichotomies. Philos Trans R Soc Lond B Biol Sci 363:1087–1104.
Zatorre RJ, Evans AC, Meyer E, Gjedde A (1992) Lateralization of phonetic and pitch discrimination in speech processing. Science 256:846–849.
Zatorre RJ, Belin P, Penhune VB (2002) Structure and function of auditory cortex: music and speech. Trends Cogn Sci 6:37–46.