Behavioral/Systems/Cognitive

# Encoding of Spectral Correlation over Time in Auditory Cortex

**Tobias Overath,**[1,2] **Sukhbinder Kumar,**[2] **Katharina von Kriegstein,**[1,2] **and Timothy D. Griffiths**[1,2]

[1]Wellcome Trust Centre for Neuroimaging, Institute of Neurology, University College London, London WC1N 3AR, United Kingdom, and [2]Newcastle Auditory Group, Medical School, Newcastle University, Newcastle-upon-Tyne NE2 4HH, United Kingdom

Natural sounds contain multiple spectral components that vary over time. The degree of variation can be characterized in terms of correlation between successive time frames of the spectrum, or as a time window within which any two frames show a minimum degree of correlation: the greater the correlation of the spectrum between successive time frames, the longer the time window. Recent studies suggest differences in the encoding of shorter and longer time windows in left and right auditory cortex, respectively. The present functional magnetic resonance imaging study assessed brain activation in response to the systematic variation of the time window in complex spectra that are more similar to natural sounds than in previous studies. The data show bilateral activity in the planum temporale and anterior superior temporal gyrus as a function of increasing time windows, as well as activity in the superior temporal sulcus that was significantly lateralized to the right. The results suggest a coexistence of hierarchical and lateralization schemes for representing increasing time windows in auditory association cortex.

*Key words:* auditory cortex; time windows; spectrotemporal correlation; fMRI; sound; speech

## Introduction

Accumulating evidence suggests that auditory perception extracts acoustic information over different time scales simultaneously. In speech, for example, phonemic and syllabic rates operate over two distinct time scales, the former on the order of tens of milliseconds, the latter over hundreds of milliseconds (Rosen, 1992). One model of speech perception, the "asymmetric sampling in time" (AST) hypothesis (Poeppel, 2003), draws on this dissociation. It posits a lateralization scheme in auditory cortex (AC) by which slower modulations (~3–6 Hz or ~150–300 ms) preferentially engage right AC, whereas fast modulations (~20–40 Hz or ~25–50 ms) preferentially engage left AC. We consider here generic mechanisms for the analysis of the temporal structure of novel sounds with a similar level of complexity to that of speech sounds.

Previous investigations (Zatorre and Belin, 2001; Boemio et al., 2005; Schönwiesner et al., 2005) manipulated the segment length within multiple-segment sounds to probe for distinct processing of different time windows. However, results in these studies differed with respect to specializations of different subareas in auditory cortex for different time windows, either within or between hemispheres. For example, Boemio et al. (2005) demonstrated sensitivity to increasing time windows in auditory association cortex (AAC), with a right-hemispheric bias in right superior temporal sulcus (STS), but no differential temporal sen-

sitivity in primary or secondary auditory cortices (PAC and SAC) as part of Heschl's gyrus (HG). However, others demonstrated sensitivity to decreasing time windows in left HG, but no differential temporal sensitivity in AAC (Zatorre and Belin, 2001; Schönwiesner et al., 2005).

Obleser et al. (2008) manipulated the spectral and temporal resolution of natural speech signals, demonstrating slight lateralization preferences in right and left AAC (specifically STS) for spectral and temporal resolution, respectively. Thus, critical yet unresolved questions relate to (1) the extent to which the analysis of different levels of temporal structure depends on PAC and SAC as opposed to AAC, and (2) the lateralization of temporal analysis within these different areas (Hickok and Poeppel, 2007; Zatorre and Gandour, 2008).

In this study we introduce a novel stimulus (see Fig. 1) based on the systematic manipulation of the degree of statistical fluctuation over time in complex acoustic spectra (for a detailed description of the stimulus see Materials and Methods; supplemental audio files of the example stimuli depicted in Fig. 1 are available at www.jneurosci.org as supplemental material). The rate of fluctuation is operationalized as the mean Pearson product-moment correlation ($r$) between amplitude spectra in adjacent time frames (Krumhansl, 1989; Krimphoff, 1993, 1994; McAdams et al., 1995; Caclin et al., 2005). Rapid modulation of the spectrum (at the phonemic rate in speech sounds) corresponds to short time windows with a minimum degree of correlation between the spectra at two time frames, even if these are not adjacent. Slow modulation of the spectrum (at the syllabic rate in speech sounds) corresponds to long time windows with a minimum degree of correlation between the spectra at any two time frames. These stimuli mimic more closely the acoustic complexity of speech and other naturally occurring sounds than those
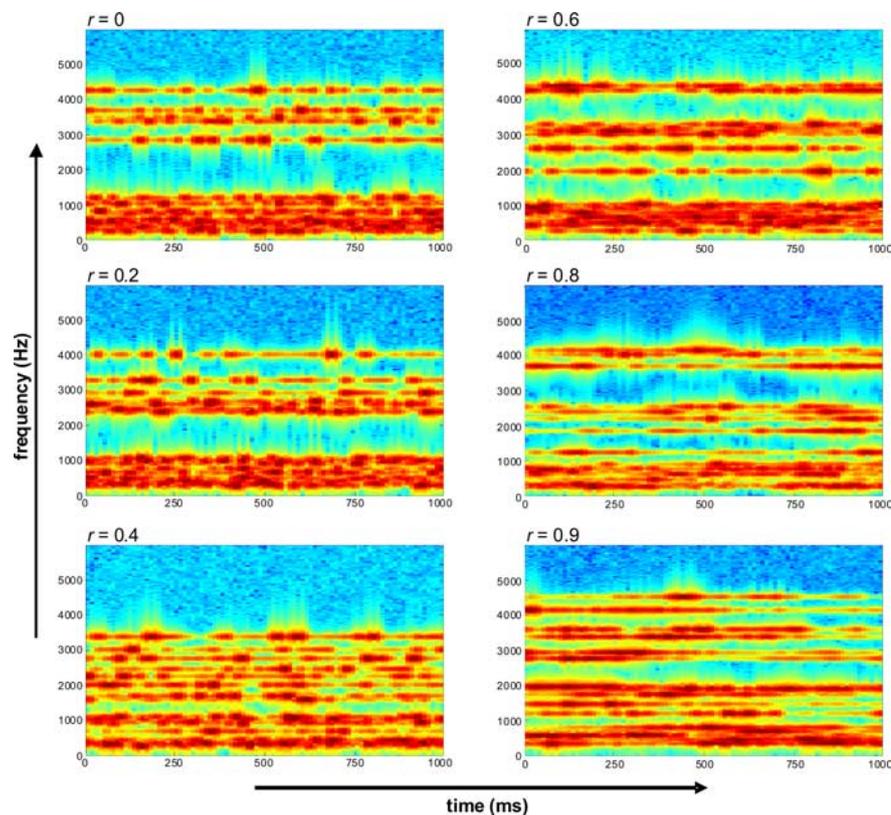
**Figure 1.** Spectrograms of representative stimuli from each level of correlation (see also supplemental sound examples of auditory stimuli, available at www.jneurosci.org as supplemental material).

$$r(x,y) = -\frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{s_x s_y},$$

where $x$ and $y$ are the amplitude (in dB) vectors over the 20 frequency components of two consecutive frames, $n$ is the number of frequencies, $s_x$ and $s_y$ represent the standard deviations of $x$ and $y$, and $\bar{x}$ and $\bar{y}$ are the arithmetic means of $x$ and $y$, respectively. The amplitude spectrum of a given sound was allowed to vary with one of six specified correlations ($r = 0, 0.2, 0.4, 0.6, 0.8, 0.9$) between the 20 ms segments.

For any value of $r$, the stimuli can be considered in terms of either correlation between adjacent time frames or in terms of a time window within which a minimum degree of correlation will exist between any two instantaneous spectra within it. The correlation between any two instantaneous spectra decays exponentially with the time frame (or lag) between the two spectra, with a rate determined by the specified correlation ($r$) between the spectra at consecutive frames. The window length is defined as the duration over which the correlation falls to a predefined lower value ($r = 0.2$, in the present case), and is calculated by the following equation:

$$\text{window\_length}(0.2) = \text{frame\_duration} \cdot \frac{\ln(0.2)}{\ln(|r|)}.$$

used in previous studies. Unlike speech, however, they allow systematic manipulation of the time windows over which correlation-controlled change in the spectrum occurs without any semantic confound, enabling the investigation of fundamental mechanisms for timing analysis.

Participants listened to stimuli with one of six levels of correlation while undergoing functional magnetic resonance imaging (fMRI). We specifically sought differences in activation as a function of increasing and decreasing correlation and the associated window length between (1) PAC and SAC versus AAC, and (2) the two hemispheres.

## Materials and Methods

*Participants.* Seventeen right-handed participants (aged 18–31, mean age = 25.35; 9 females) with normal hearing and no history of audiological or neurological disorders provided written consent before the study. The study was approved by the Institute of Neurology Ethics Committee, London, UK.

*Stimuli.* All stimuli (Fig. 1) were created digitally in the frequency domain using Matlab software (MathWorks) at a sampling rate of 44.1 kHz and 16 bit resolution. Each sound consisted of 20 sinusoids randomly chosen from a pool of 101 logarithmically spaced frequencies between 246 and 4435 Hz. The particular parameters were chosen so as to approximate respective features in naturally occurring sounds, which typically have complex spectra with multiple frequencies present. The bandpass (246–4435 Hz) covers the acoustic range that is most important for human auditory perception, and the number of frequencies within this pool (101 frequencies) are a result of this range. The amplitude spectrum was defined in 20 ms frames such that the correlation from one frame to the next was operationalized as the Pearson product moment correlation $r$:

Figure 2 (inset) shows the relationship between the window length and the correlation ($r$) when the frame duration is 20 ms: the window length corresponding to values of $r$ between 0.2 and 0.9 varies between 20 ms and 305 ms, encompassing windows relevant to phonemic and syllabic processing, respectively (Rosen, 1992).

Linear spline interpolation was applied to amplitude transitions between frames to avoid sudden amplitude jumps. Importantly, the mean amplitude (65 dB) and standard deviation (SD = 15) were identical for each frequency component in a given sound and across correlation levels. Each sound had a rise and fall time of 20 ms.

*Experimental design.* Before the experiment in the MRI scanner, participants were familiarized with the stimuli and then performed two-interval two-alternative forced choice psychophysics with $r = 0$ as reference sounds. Stimuli were 2 s long and were different exemplars from the ones subsequently used in the scanner. Psychophysics ensured that participants were able to distinguish a highly correlated sound from the reference sound, and they needed to reach at least 90% correct performance for the strongest correlation ($r = 0.9$) to be included in the fMRI study.

Stimuli in the scanner were of different durations (1, 2, 3, or 4 s) and separated by a mean interstimulus interval (ISI) of 2 s. (range: 1.5–2.5 s) as well as occasional silence trials of 6 s duration (20 per session). Stimuli were presented in a pseudorandom order, with 20 exemplars for each level per session (80 stimuli per level in total, amounting to a total presentation time of 200 s per level). Participants performed a stimulus-irrelevant task by pressing a button after each sound.

Stimuli were presented via NordicNeuroLab electrostatic headphones at 80 dB sound pressure level (SPL) using Cogent software (http://www.vislab.ucl.ac.uk/Cogent).

*Image acquisition.* Gradient weighted echo planar images (EPI) were acquired on a 3 Tesla Siemens Allegra system, using a continuous imaging design with 42 contiguous slices per volume (time to repeat/time to echo, 2730/30 ms). The volume was tilted forward such that slices were parallel to the superior temporal plane. Participants completed four ses-

sions of 250 volumes each, resulting in a total of 1000 volumes. To correct for geometric distortions in the EPI due to B0 field variations, Siemens fieldmaps were acquired for each subject, usually after the second session (Hutton et al., 2002; Cusack et al., 2003). A structural T1 weighted scan was acquired for each participant (Deichmann et al., 2004).

*Image analysis.* Imaging data were analyzed using Statistical Parametric Mapping software (SPM5, http://www.fil.ion.ucl.ac.uk/spm). The first four volumes in each session were discarded to control for saturation effects. The resulting 984 volumes were realigned to the first volume and unwarped using the fieldmap parameters, spatially normalized to stereotactic space (Friston et al., 1995a) and smoothed with an isotropic Gaussian kernel of 8 mm full-width-at-half-maximum (FWHM). Contrast values at the subject level probing for an effect of time window length over the six levels were derived from the standard exponential decay function (see Fig. 2, inset formula) and then mean centered on zero to yield $[-2.94, -2.32, -1.85, -0.98, 1.54, 6.55]$. Statistical analysis at the group level used a random-effects model within the context of the general linear model (Friston et al., 1995b), and data were thresholded at $p < 0.001$ (uncorrected for multiple comparisons across the brain) for areas where we had an a priori hypothesis, i.e., auditory cortex. Where the results survived a more conservative threshold of $p < 0.05$ (family-wise error corrected for multiple comparisons across the brain), we report results at this threshold.

For the test of lateralization, two sets of images were created: both a set of "flipped" left-right unwarped images as well as the original unwarped images were normalized to a symmetrical template so as to enable a direct comparison between the activations in the left and right AC. Note that the resulting symmetrical stereotactic space will differ slightly from MNI stereotactic space. These original and flipped normalized scans were smoothed with an 8 mm FWHM smoothing kernel, as above. Both original and flipped scans were then combined in one design to enable a direct comparison. Statistical analysis at the group level was thresholded at $p < 0.001$ (uncorrected for multiple comparisons across the brain).

To compare in detail the response in subareas of auditory cortex as a function of spectrotemporal correlation, we identified local maxima coordinates based on the main contrast of spectrotemporal correlation [for planum temporale (PT), anterior superior temporal gyrus (aSTG), and STS], and based on a [sound–silence] contrast for left and right HG that are most similar to central HG or SAC (Morosan et al., 2001; Rademacher et al., 2001; Patterson et al., 2002). We then extracted the parameter estimates of the BOLD signal at these coordinates.

## Results

An analysis was performed to seek areas where the activity increased or decreased as a function of correlation and the associated window length (see Materials and Methods). The results show bilateral activity in AAC as a function of increasing correlation or temporal window, in particular in PT and aSTG, while also extending into right STS [Fig. 2, see also Table 1 for coordinates of local maxima and supplemental materials (available at www.jneurosci.org)] for further discussion of the precise shape of the relationship between BOLD signal and correlation $r$ in left and right AAC). We formally tested whether this effect arises in and is specific to AAC in PT and aSTG and is not already present in HG (see also Fig. 2) by extracting the BOLD signal (see Materials and Methods) in central HG, which is most similar to SAC (Morosan et al., 2001; Rademacher et al., 2001; Patterson et al., 2002), and the association areas that showed an increase in activity as a function of correlation. Two separate (for PT and aSTG) repeated measures ANOVAs with factors 2 Hemisphere (left, right) × 2 Area (HG, [PT or aSTG]) × 6 Correlation level (1–6) demonstrated an Area × Correlation level interaction ($F_{(5,80)} = 8.28, p < 0.001$ for PT; and $F_{(5,80)} = 5.19, p < 0.01$ for aSTG).

To compare directly the response in left and right auditory cortices, we performed a formal test of lateralization by "flipping" and normalizing the functional scans to a symmetrical template (see Materials and Methods). Activity in PT and aSTG did not

differ between left and right hemispheres. However, right STS showed significantly stronger activation as a function of increasing correlation than its left hemisphere homolog (Fig. 3; Table 1). A repeated-measures ANOVA with 2 Area (left STS, right STS) × 6 Correlation level (1–6) as factors revealed a significant interaction ($F_{(5,80)} = 2.33, p = 0.05$). That is, while PT and aSTG in both hemispheres are equally involved in processing longer spectrotemporal correlation over time, the data suggest a right-lateralized preference in STS.

We found no evidence for an effect of decreasing correlation; that is, no area showed a signal increase as the time window associated with each level of correlation decreased. Even lowering the statistical threshold to a very lenient $p = 0.1$ (uncorrected for multiple comparisons) did not yield any activation in the auditory system. Consequently, there was no detectable lateralization as a function of decreasing correlation.

## Discussion

We systematically varied the spectrotemporal correlation in complex sounds and demonstrated an increase in activation in AAC as a function of spectral correlation over time (or equivalently as a function of lengthening temporal window). PT and aSTG showed a bilateral increase in activity with increasing correlation and we show that this relationship arises in AAC and is not already present in Heschl's gyrus (i.e., in PAC or SAC). Furthermore, activity along the upper bank of right STS increased to a greater extent than left STS as a function of increasing correlation. We did not observe any areas that showed an increase in activity as a function of decreasing correlation over time (shorter time windows).

The stimuli in the current study were based on complex spectra with multiple components which varied over time in statistically controlled ways that are similar to ethological sounds including speech. In contrast, previous neurophysiological studies of temporal analysis in animal cortex have generally used more deterministic stimuli including sinusoidal amplitude modulation of narrow-band stimuli or noise (Joris et al., 2004) (but see Malone et al., 2007). Neurophysiological studies of amplitude modulation in (mainly primary) auditory cortex in humans (Liégeois-Chauvel et al., 2004) and mammals (Joris et al., 2004) show preferred responses to rates of <20 Hz, corresponding to temporal windows at the level of tens to hundreds of milliseconds, as used in the present study. Several human imaging studies have used more complex types of temporal modulation (Zatorre and Belin, 2001; Boemio et al., 2005; Schönwiesner et al., 2005), but none have controlled the change in complex spectra from one moment to the next as in the present study.

We explicitly tested the contribution of different areas of auditory cortex and demonstrated different response profiles across the six levels of correlation between HG on the one hand and AAC on the other. HG did not differentiate between the experimental levels, while AAC, with maxima in PT and aSTG, displayed a systematic BOLD signal increase as a function of spectral correlation. Previous models have tended to emphasize differences in temporal analysis between hemispheres, rather than differences between the specific areas of auditory cortex within hemispheres or differences between lateralization in different areas. Human anatomical (Morosan et al., 2001; Rademacher et al., 2001) and functional imaging studies (e.g., Patterson et al., 2002) have demonstrated one primary and two secondary areas in Heschl's Gyrus that might correspond to "core" areas in macaque, as opposed to human homologues of belt areas of AAC in the planum temporale (PT) and superior temporal gyrus STG (Hackett,
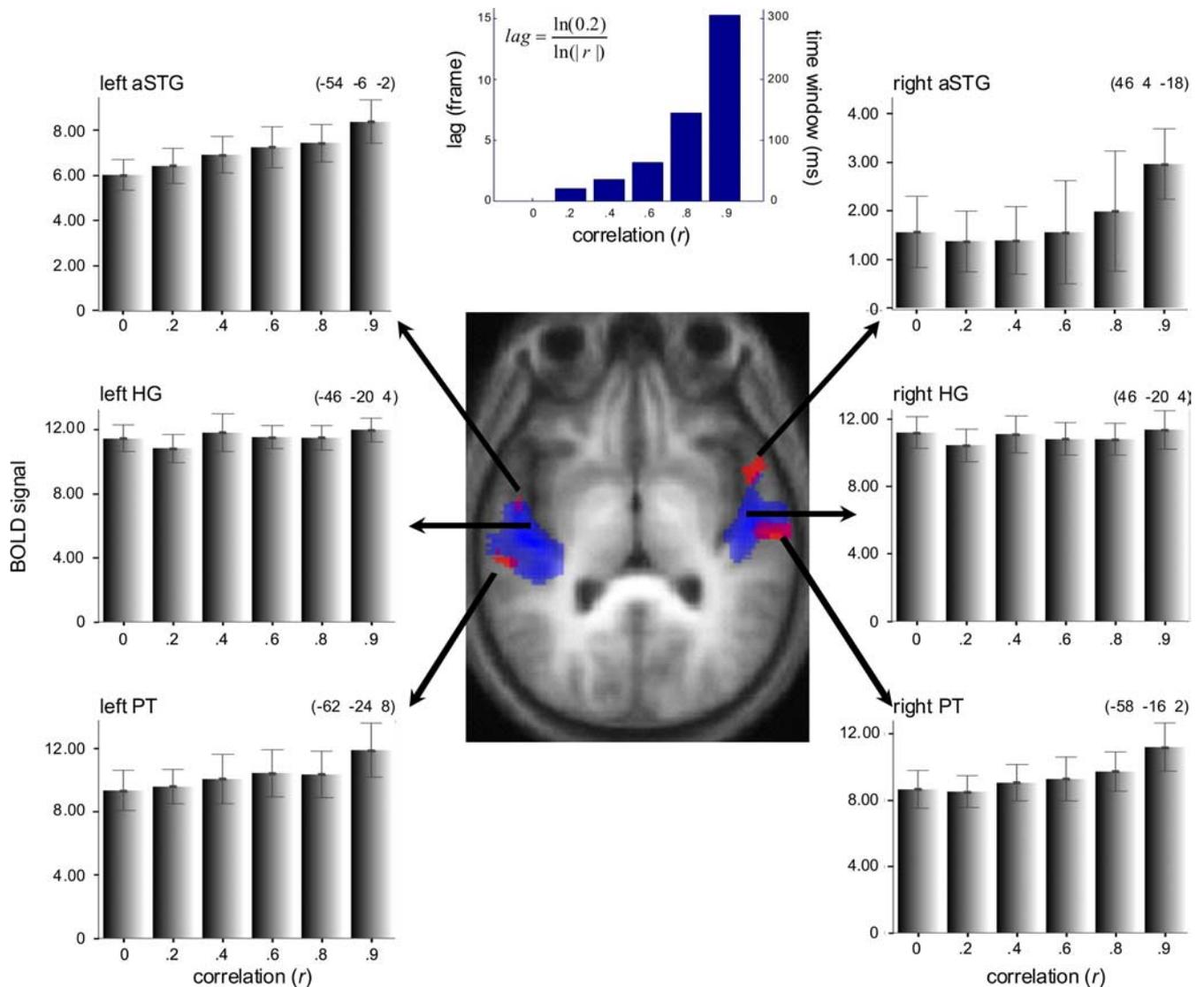
**Figure 2.** Areas increasing in activity as a function of spectrotemporal correlation (red) and areas responding to sound in general (blue). Results are rendered on a tilted (pitch = −0.5) section of the normalized average structural along STG and thresholded at $p < 0.05$ (familywise error corrected). The bar plots at the sides show the signal at the respective coordinates for the six levels of correlation (± 95% confidence interval). The top figure displays the average lag (in 20 ms frames) and associated time window length (in ms) for which there exists a correlation $r > 0.2$ for the six levels of correlation, as determined by the exponential decay function (inset formula).

**Table 1. MNI coordinates of local maxima ($p < 0.05$, familywise error corrected) in PT and aSTG as a function of increasing time window correlation and coordinates of local maxima ($p < 0.001$, uncorrected) in right STS for the lateralization test**

| | | Left | | | | Right | | | |
|---|---|---|---|---|---|---|---|---|---|
| Contrast | Area | x | y | z | t value | x | y | z | t value |
| Window increase | PT | −62 | −24 | 8 | 10.52 | 58 | −16 | 2 | 13.49 |
| | | | | | | 66 | −16 | 0 | 12.83 |
| | aSTG | −50 | −6 | −14 | 11.1 | 46 | 4 | −18 | 9.91 |
| | | −54 | −6 | −2 | 8.71 | 56 | 10 | −12 | 9.54 |
| Lateralization | STS | | | | | 66 | −22 | −8 | 5.1 |
| | | | | | | 56 | −44 | 6 | 4.56 |
| | | | | | | 54 | −10 | −16 | 5.34 |
| | | | | | | 48 | 12 | −24 | 3.9 |

Note that the coordinates for STS are only approximations, since they were normalized to a symmetrical template.

2007). Areas of AAC in the superior temporal sulcus may correspond to parabelt in the macaque. While the homology with macaque schemes is still being explored, it is clear that there is an extensive functional architecture for auditory analysis that might allow different subspecializations for various types of temporal

analysis between areas and between the hemispheres. The present study demonstrates subspecializations for auditory analysis between different areas, and does not support any simple model based on similar temporal analysis in all the auditory cortical areas on either side.
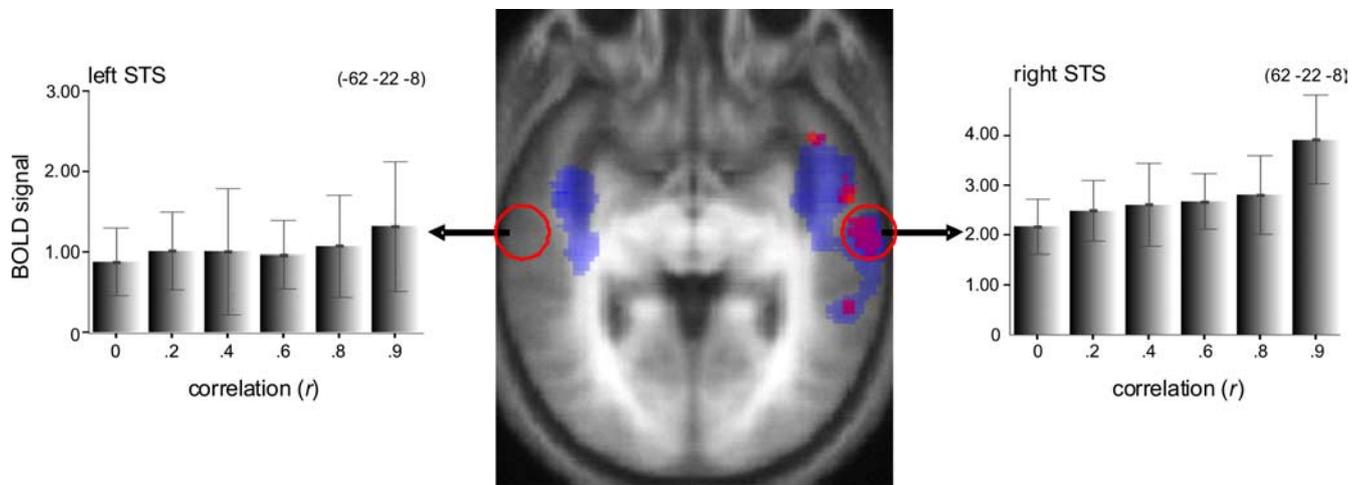
**Figure 3.** Areas showing a significantly stronger increase in activity in right STS than left STS (red) together with areas that show an increase as a function of correlation (blue). Results are rendered on a tilted (pitch = −0.5) section of the symmetrical normalized average structural along STS and thresholded at $p < 0.001$ (uncorrected). The bar plots at the sides show the signal at the respective coordinates for the six levels of correlation (± 95% confidence interval).

Using BOLD as a measure of local ensemble activity, the data did not show a preference for short time windows (at the level of tens of milliseconds) in any area of auditory cortex, even when the statistical threshold was lowered substantially. A potential explanation for this might be the existence of different neural coding schemes for slow versus fast temporal modulations to which the BOLD signal might not be as sensitive. For example, Lu et al. (2001) (see also Wang et al., 2003) have demonstrated that slow temporal modulations are encoded explicitly via synchronized discharge rates, whereas fast modulations are encoded implicitly via nonsynchronized discharges. There is further evidence that neural synchronization in the gamma frequency range is tightly coupled to the hemodynamic response in cortex (Niessing et al., 2005).

However, some studies have reported signal increases as a function of increasing rates of temporal modulation. Specifically, Zatorre and Belin (2001) and Schönwiesner et al. (2005) demonstrated increased activity in primary and secondary auditory cortex with increasing rate of sound fluctuation (see also Jamison et al., 2006). Zatorre and Belin altered the fluctuation rate of two sinusoidal components (500 and 1000 Hz) and thus arguably used a substantially different stimulus compared with the present study; however, the broadband stimuli in Schönwiesner et al. (2005) are similar in acoustic complexity to those used in the present study, although not controlling changes in the spectral shape from one moment to the next as here. The different results between studies pose a paradox, since they have used a similar experimental manipulation (sound segment length) and similar temporal window length to test a left-lateralization for increasing temporal fluctuations posited by both AST (Poeppel, 2003) and spectro-temporal trade-off (Zatorre et al., 2002) theories.

A promising recent approach (Giraud et al., 2007) combined fMRI and EEG recordings of spontaneous spectral power (in the absence of any experimental acoustic stimulation, but in the presence of scanner noise) to test the AST hypothesis and found activity in left and right HG (but not AAC) that correlated with fast (∼28–40 Hz) and slow (∼3–6 Hz) neural oscillations, respectively. While these findings somewhat contradict the precise anatomical locations of the earlier findings of Boemio et al. (2005) for longer temporal windows, they might nevertheless offer a bridge in that they show a left-lateralization for fast tem-

poral modulations (as posited by both AST and spectro-temporal trade-off theories) and a right-lateralization for slower temporal modulations (as posited by AST). However, comparisons between studies of spontaneous activity in the absence of experimental acoustic input and the temporal structure of stimuli producing the greatest regional activity need to be made with caution. A recent study (Obleser et al., 2008) that manipulated the spectral and temporal resolution within degraded speech stimuli and found corresponding differential engagement of right and left auditory association cortex, respectively, points to the importance of the interplay of both temporal and spectral information.

A convincing explanation for the divergence of results between the previous studies despite similar experimental manipulations has yet to be provided (Hickok and Poeppel, 2007; Zatorre and Gandour, 2008).

The present study has demonstrated the analysis of longer temporal windows in the syllabic range that is bilateral in AAC in STG and right lateralized in STS. In Boemio et al. (2005), the analysis of longer time segments was similarly right lateralized and involved STS. An open question remains as to why longer temporal windows, which are important for syllabic information and speech intelligibility (Greenberg et al., 2003; Luo and Poeppel, 2007), should be lateralized toward right AC, as posited by the AST hypothesis (Poeppel, 2003). For example, Scott et al. (2000) (Narain et al., 2003) argue that speech intelligibility engages a left-lateralized network in the temporal lobe. In contrast, the present data, as well as those of Boemio et al. (2005), revealed a significant right-lateralization in STS for increasing time windows.

Furthermore, although occurring in "higher" association cortex, the effect is present although participants in the current study performed a stimulus-irrelevant task (Boemio et al., 2005, used no task), and can thus be argued to be an obligatory correlate of perception. We would point out that this level of temporal analysis is relevant to a variety of sounds including speech, but is not specific to speech.

The notion of increasing temporal integration windows in higher cortex is not specific to the analysis of acoustic signals, but has also been demonstrated in the visual (Hasson et al., 2008) and motor (Schubotz, 2007) systems, suggesting a general processing

scheme within cortex (Kiebel et al., 2008). A possible neural mechanism in auditory cortex might be the differential existence of transient versus sustained responses in primary and association cortex, respectively (Seifritz et al., 2002). Here we hypothesize that relative lateralization preferences are a secondary phenomenon to a hierarchical processing scheme in which the length of analysis time windows increases as one progresses along the hierarchy in auditory cortex.

The present study highlights the potency of parametrically varying statistical properties of complex acoustic stimuli to investigate systematically principles of processing in auditory cortex (Nelken and Chechik, 2007; Overath et al., 2007). We present a novel stimulus with characteristics that vary in a similar manner to naturally occurring sounds including speech and demonstrate a network comprising auditory association cortex that plays a crucial role in tracking spectral correlation over different time scales.

## References

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

Caclin A, McAdams S, Smith BK, Winsberg S (2005) Acoustic correlates of timbre space dimensions: a confirmatory study using synthetic tones. J Acoust Soc Am 118:471–482.

Cusack R, Brett M, Osswald K (2003) An evaluation of the use of magnetic field maps to undistort echo-planar images. Neuroimage 18:127–142.

Deichmann R, Schwarzbauer C, Turner R (2004) Optimization of the 3D MDEFT sequence for anatomical brain imaging: technical implications at 1.5 T and 3 T. Neuroimage 21:757–767.

Friston KJ, Ashburner J, Frith CD, Poline JB, Heather JD, Frackowiak RS (1995a) Spatial registration and normalisation of images. Hum Brain Mapp 2:165–189.

Friston KJ, Holmes AP, Worsley KJ, Poline JB, Frith CD, Frackowiak RS (1995b) Statistical parametric maps in functional imaging: a general linear approach. Hum Brain Mapp 2:189–210.

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.

Greenberg S, Carvey H, Hitchcock L, Chang S (2003) Temporal properties of spontaneous speech—a syllable-centric perspective. J Phonetics 31:465–485.

Hackett TA (2007) Organization and correspondence of the auditory cortex of humans and nonhuman primates. In: Evolution of nervous systems (Kaas JH, ed), pp 109–119. Oxford: Elsevier.

Hasson U, Yang E, Vallines I, Heeger DJ, Rubin N (2008) A hierarchy of temporal receptive windows in human cortex. J Neurosci 28:2539–2550.

Hickok G, Poeppel D (2007) The cortical organization of speech processing. Nat Rev Neurosci 8:393–402.

Hutton C, Bork A, Josephs O, Deichmann R, Ashburner J, Turner R (2002) Image distortion correction in fMRI: a quantitative evaluation. Neuroimage 16:217–240.

Jamison HL, Watkins KE, Bishop DV, Matthews PM (2006) Hemispheric specialization for processing auditory nonspeech stimuli. Cereb Cortex 16:1266–1275.

Joris PX, Schreiner CE, Rees A (2004) Neural processing of amplitude-modulated sounds. Physiol Rev 84:541–577.

Kiebel SJ, Daunizeau J, Friston KJ (2008) A hierarchy of time-scales and the brain. PLoS Comp Biol 14:e1000209.

Krimphoff J (1993) Analyse acoustique et perception du timbre. DEA thesis, Université du Maine, Le Mans.

Krimphoff J, McAdams S, Winsberg S (1994) Caractérisation du timbre des sons complex. II Analyses acoustiques et quantification psychoacoustique [Characterization of the timbre of complex sounds II Acoustic analyses and psychophysical quantification]. J Phys (Paris) 4:625–628.

Krumhansl CL (1989) Why is musical timbre so hard to understand? In: Structure and perception of electroacoustic sound and music (Nielzen S, Olsen O, eds), pp 43–53. Amsterdam: Excerpta Medica.

Liégeois-Chauvel C, Lorenzi C, Trébuchon A, Régis J, Chauvel P (2004) Temporal envelope processing in the human left and right auditory cortices. Cereb Cortex 14:731–740.

Lu T, Liang L, Wang X (2001) Temporal and rate representations of time-varying signals in the auditory cortex of awake primates. Nat Neurosci 4:1131–1138.

Luo H, Poeppel D (2007) Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. Neuron 54:1001–1010.

Malone BJ, Scott BH, Semple MN (2007) Dynamic amplitude coding in the auditory cortex of awake rhesus macaques. J Neurophysiol 98:1451–1474.

McAdams S, Winsberg S, Donnadieu S, De Soete G, Krimphoff J (1995) Perceptual scaling of synthesized musical timbres: common dimensions, specificities, and latent subject classes. Psychol Res 58:177–192.

Morosan P, Rademacher J, Schleicher A, Amunts K, Schormann T, Zilles K (2001) Human primary auditory cortex: cytoarchitectonic subdivisions and mapping into a spatial reference system. Neuroimage 13:684–701.

Narain C, Scott SK, Wise RJ, Rosen S, Leff A, Iversen SD, Matthews PM (2003) Defining a left-lateralized response specific to intelligible speech using fMRI. Cereb Cortex 13:1362–1368.

Nelken I, Chechik G (2007) Information theory in auditory research. Hear Res 229:94–105.

Niessing J, Ebisch B, Schmidt KE, Niessing M, Singer W, Galuske RA (2005) Hemodynamic signals correlate tightly with synchronized gamma oscillations. Science 309:948–951.

Obleser J, Eisner F, Kotz SA (2008) Bilateral speech comprehension reflects differential sensitivity to spectral and temporal features. J Neurosci 28:8116–8123.

Overath T, Cusack R, Kumar S, von Kriegstein K, Warren JD, Grube M, Carlyon RP, Griffiths TD (2007) An information theoretic characterisation of auditory encoding. PLoS Biol 5:e288.

Patterson RD, Uppenkamp S, Johnsrude IS, Griffiths TD (2002) The processing of temporal pitch and melody information in auditory cortex. Neuron 36:767–776.

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.

Rosen S (1992) Temporal information in speech: acoustic, auditory and linguistic aspects. Philos Trans R Soc Lond B Biol Sci 336:367–373.

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.

Schubotz RI (2007) Prediction of external events with our motor system: towards a new framework. Trends Cogn Sci 11:211–218.

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.

Seifritz E, Esposito F, Hennel F, Mustovic H, Neuhoff JG, Bilecen D, Tedeschi G, Scheffler K, Di Salle F (2002) Spatiotemporal pattern of neural processing in the human auditory cortex. Science 297:1706–1708.

Wang X, Liu T, Liang L (2003) Cortical processing of temporal modulations. Speech Commun 41:107–121.

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, Belin P, Penhune VB (2002) Structure and function of auditory cortex: music and speech. Trends Cogn Sci 6:37–46.