

Behavioral/Systems/Cognitive

Cortical Mechanisms for the Segregation and Representation of Acoustic Textures

Tobias Overath,^{1,2,3} Sukhbinder Kumar,² Lauren Stewart,⁴ Katharina von Kriegstein,^{1,2,5} Rhodri Cusack,⁶ Adrian Rees,² and Timothy D. Griffiths^{1,2}

¹Wellcome Trust Centre for Neuroimaging, Institute of Neurology, University College London, London WC1N 3AR, United Kingdom, ²Newcastle Auditory Group, Medical School, Newcastle University, Newcastle-upon-Tyne NE2 4HH, United Kingdom, ³Department of Psychology, New York University, New York, New York 10003, ⁴Department of Psychology, Goldsmiths, University of London, New Cross, London SE14 6NW, United Kingdom, ⁵Max Planck Institute for Human Cognitive and Brain Sciences, 04103 Leipzig, Germany, and ⁶Medical Research Council Cognition and Brain Sciences Unit, University of Cambridge, Cambridge CB2 7EF, United Kingdom

Auditory object analysis requires two fundamental perceptual processes: the definition of the boundaries between objects, and the abstraction and maintenance of an object's characteristic features. Although it is intuitive to assume that the detection of the discontinuities at an object's boundaries precedes the subsequent precise representation of the object, the specific underlying cortical mechanisms for segregating and representing auditory objects within the auditory scene are unknown. We investigated the cortical bases of these two processes for one type of auditory object, an "acoustic texture," composed of multiple frequency-modulated ramps. In these stimuli, we independently manipulated the statistical rules governing (1) the frequency–time space within individual textures (comprising ramps with a given spectrotemporal coherence) and (2) the boundaries between textures (adjacent textures with different spectrotemporal coherences). Using functional magnetic resonance imaging, we show mechanisms defining boundaries between textures with different coherences in primary and association auditory cortices, whereas texture coherence is represented only in association cortex. Furthermore, participants' superior detection of boundaries across which texture coherence increased (as opposed to decreased) was reflected in a greater neural response in auditory association cortex at these boundaries. The results suggest a hierarchical mechanism for processing acoustic textures that is relevant to auditory object analysis: boundaries between objects are first detected as a change in statistical rules over frequency–time space, before a representation that corresponds to the characteristics of the perceived object is formed.

Introduction

Natural sounds are both spectrally as well as temporally complex, and a fundamental question in auditory perception is how the brain represents and segregates distinct auditory objects that contain many individual spectrotemporally varying components. Such auditory object analysis requires at least two fundamental perceptual processes. The first is the detection of boundaries between adjacent objects, which necessitates mechanisms that identify changes in the statistical rules governing object regions in frequency–time space (Kubovy and Van Valkenburg, 2001; Chait et al., 2007, 2008). The second is the abstraction and maintenance

of an object's characteristic features, while ignoring local stochastic variation within an object (Griffiths and Warren, 2004; Nelken, 2004).

Although it is intuitive to assume that the detection of statistical changes at object boundaries precedes the subsequent precise representation of the object (Ohl et al., 2001; Chait et al., 2007; 2008; Scholte et al., 2008), the specific underlying cortical mechanisms for segregating and representing auditory objects within the auditory scene have not been directly addressed. Zatorre and colleagues (2004) demonstrated a parametric increase in activity within right superior temporal sulcus (STS) as a function of object distinctiveness. However, whether this effect was due to object distinctiveness as such or due to a change percept between objects is unclear. That is, as the distinctiveness between objects increased, participants were also more likely to hear a change at object boundaries.

Other studies focused on the neural correlates of boundary or "auditory edge" detection without investigating in detail processes necessary for object formation (Chait et al., 2007; 2008). Schönwiesner et al. (2007) investigated the perception of different levels of changes in acoustic duration. They found a cortical hierarchy for processing duration changes as indicated by three distinct stages: an initial change detection mechanism in primary auditory cortex, followed by a more detailed analysis in associa-

Received Oct. 30, 2009; accepted Dec. 2, 2009.

This work was funded by the Wellcome Trust, UK. We thank two anonymous reviewers for their helpful comments. T.O. designed and performed the experiment, analyzed the data, and wrote this manuscript. S.K. created the stimulus, designed the experiment, and wrote this manuscript. L.S. performed pilot studies with the stimulus. K.v.K. helped with data acquisition and analysis and with writing this manuscript. R.C. contributed to data analysis and to writing this manuscript. A.R. contributed to stimulus design and wrote this manuscript. T.D.G. created the stimulus, designed the experiment, and wrote this manuscript.

Correspondence should be addressed to either of the following: Dr. Tobias Overath, Department of Psychology, New York University, 6 Washington Place, New York, NY 10003, E-mail: t.overath@nyu.edu; or Dr. Timothy D. Griffiths, Newcastle Auditory Group, Medical School, Newcastle University, Framlington Place, Newcastle-upon-Tyne NE2 4HH, UK, E-mail: t.d.griffiths@newcastle.ac.uk.

DOI:10.1523/JNEUROSCI.5378-09.2010

Copyright © 2010 the authors 0270-6474/10/302070-07\$15.00/0

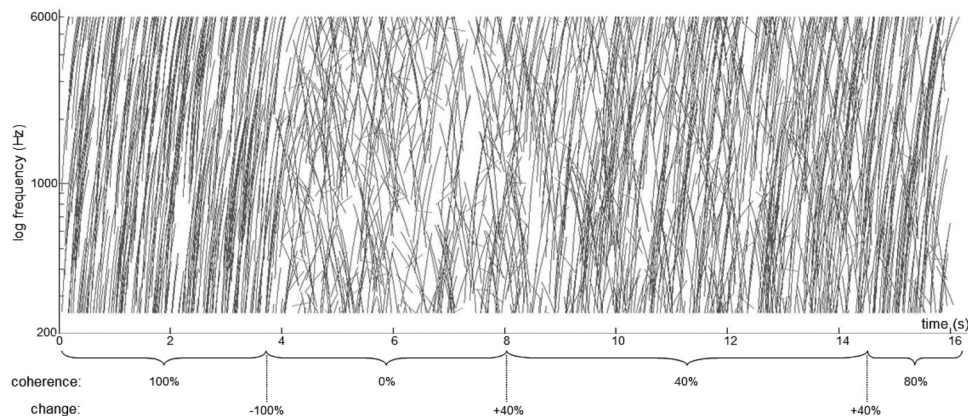


Figure 1. Auditory stimulus. Example of a block of sound with four spectrotemporal coherence segments showing absolute coherence values for each segment and the corresponding change in coherence between the segments.

tion cortex and attentional mechanisms originating in frontal cortex.

The present study used a form of spectrotemporal coherence to create object regions and object boundaries in frequency–time space. The stimulus, an “acoustic texture,” was based on randomly distributed linear frequency-modulated (FM) ramps with varying trajectories, where the overall coherence between ramps was controlled. Conceptually, the stimulus is similar to the coherent visual motion paradigm (Newsome and Paré, 1988; Rees et al., 2000; Braddick et al., 2001). In both, the coherence of constituent elements can be parametrically controlled. The analysis of acoustic textures comprising different spectrotemporal coherences requires perceptual mechanisms that can assess common statistical properties of the stimulus, regardless of local stochastic variation within an object, and detect transitions when these properties change.

Thus, this stimulus enables a direct and orthogonal assessment of the neural correlates of the following: (1) detecting boundaries between acoustic textures with different spectrotemporal coherences; and (2) representing spectrotemporal coherence within a texture. We hypothesized that the detection of a change in spectrotemporal coherence between textures would engage auditory areas including primary cortex (Schönwiesner et al., 2007), whereas spectrotemporal coherence within textures would be encoded in higher-level auditory areas only (Zatorre et al., 2004).

Materials and Methods

Participants. Twenty-three right-handed participants (aged 18–31; mean age, 25.04; 12 females) with normal hearing and no history of audiological or neurological disorders provided written consent before the study began. The study was approved by the Institute of Neurology Ethics Committee (London, UK).

Stimuli. The acoustic texture stimulus was based on randomly distributed linear FM ramps with varying trajectories (Fig. 1). The percentage of coherent spectrotemporal modulation, i.e., the proportion of ramps with identical direction (slope–sign) and trajectory (slope–value), can be systematically controlled, creating acoustic textures with different levels of spectrotemporal coherence. Boundaries were created and their magnitude varied by juxtaposing acoustic textures of different coherence levels. In the example given in Figure 1, a 3.5 s stimulus segment with 100% spectrotemporal coherence (all ramps move upwards and with the same trajectory) is followed by a 4.5 s stimulus segment of 0% coherence (ramps move in different directions and with different trajectories) and so forth (Fig. 1; see also supplemental soundfile S1, available at www.jneurosci.org as supplemental material). The associated change in coher-

ence at the boundaries between acoustic textures with different spectrotemporal coherences is also shown.

All stimuli were created digitally in the frequency domain using MATLAB 6.5 software (The MathWorks) at a sampling frequency of 44.1 kHz and 16 bit resolution. Stimuli consisted of a dense texture of linear FM ramps; each ramp had a duration of 300 ms and started at a random time and frequency (passband, 250–6000 Hz), with a density of 80 glides per second, roughly equaling one ramp per critical band (see supplemental Fig. S2, available at www.jneurosci.org as supplemental material). For ramps that extended beyond the passband, i.e., went <250 Hz or >6000 Hz, we implemented a wraparound such that the ramps “continued” at the other extreme of the frequency band, i.e., at 6000 or 250 Hz, respectively. Stimuli differed in terms of the coherent movement of the ramps: six different coherence conditions were created, where the percentage of ramps moving in the same direction for a given sound segment was systematically varied from 0% coherence to 100% coherence in 20% increments. Thus, for a given sound segment with 40% coherence, 40% of the ramps increased (or decreased) in frequency with an excursion traversing 2.5 octaves per 300 ms; the direction and excursion of the remaining 60% of the FM ramps were randomized. Crucially, the only difference between the six levels is the degree of coherence or “common fate” of the ramps; the total number of ramps and the number of ramps in a critical band, as well as the passband of each stimulus, did not differ systematically across the levels (supplemental Fig. S2, available at www.jneurosci.org as supplemental material).

Experimental design. Before scanning, participants were familiarized with and trained on the stimuli and then performed two-interval, two-alternative forced choice (2I2AFC) psychophysics distinguishing the nonrandom against a random reference (0% coherence) sound. Stimuli were 2 s long and the direction of the FM glides (up vs down) was counterbalanced. There were 30 pairs for each of the six levels (0–100% coherence in 20% steps). Participants had to reach at least 90% correct performance for the last level (100% coherence) to be included in the functional magnetic resonance imaging study.

During the scanning session, stimuli were presented in blocks of sound with an average duration of 16 s (range: 11 to 18 s). The blocks contained four contiguous segments with a given absolute spectrotemporal coherence (0, 20, 40, 60, 80, or 100%). Within a block, the direction (up vs down) of the coherent ramps was maintained. The length of the segments varied (1.5, 3, 3.5, 4.5, 5, or 6.5 s) and was randomized within a block. Thus, a given block might have, for example, [100% 0% 40% 80%] contiguous coherence segments with durations of [3.5 4.5 6.5 1.5] seconds. The associated change in coherence between the four segments within this block of sound is [–100% +40% +40%]. Note that a +40% change in coherence, for example, can be obtained in a number of ways by suitably arranging adjacent pairs of acoustic textures with certain absolute coherence levels (0 to 40% and 40 to 80%, in the example in Fig. 1). Stimuli were presented in one of six pseudorandom permutations

that orthogonalized absolute coherence and change in coherence (average correlation between absolute coherence and change in coherence across the six permutations: $r = 0.06$, $p > 0.1$).

The task of participants was to detect a change in coherence within the block, regardless of whether that change was from less coherent to more coherent or vice versa. Participants were required to press a button whenever they heard such a change and were instructed that the frequency of perceptual changes within one block likely ranged from no perceptual change (e.g., a block consisting of [0% 20% 40% 20%] coherence segments, because here the changes are likely to be too small to be detected) to a maximum of three changes (e.g., a sound block consisting of [0% 100% 20% 80%] segments). Sound blocks were separated by a silent period of 6 s, in which participants were told to relax.

In each of three experimental scanning sessions, each coherence level was presented 30 times, amounting to a total of 7.2 min presentation time per coherence level. The number of times each of the six different levels of change in coherence (regardless of their direction) occurred can, consequently, not be perfectly balanced; however, permutations were created such that the change that occurred most often occurred less than three times as often as the change that occurred least frequently.

Stimuli were presented via NordicNeuroLab electrostatic headphones at a sound pressure level of 85 dB. Participants saw a cross at the center of the screen and were asked to look at this cross during the experiment.

Behavioral data analysis. Participants' button presses were recorded and analyzed with respect to the onset of each segment within a block. Responses were only counted if they occurred within 3 s after the onset of a segment (and within 1.5 s after the onset of the shortest segments). The average percentage correct response was then computed by comparing the number of responses to a given change in spectrotemporal coherence to the actual number of those changes. "Responses" to 0% changes served as a "false alarm" chance baseline.

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 and centered on the superior temporal gyrus. Participants completed three sessions of 372 volumes each, resulting in a total of 1116 volumes. To correct for geometric distortions in the EPI due to B0 field variations, Siemens field maps 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 1104 volumes were realigned to the first volume and unwrapped using the field map 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). Statistical analysis used a random-effects model within the context of the general linear model (Friston et al., 1995b). A region of interest (ROI) in auditory cortex was derived from a functional contrast that modeled the BOLD response to the onset of each sound block (see below); the ROI was based on a significance threshold of $p < 0.01$ [corrected for false discovery rate (FDR)] (Genovese et al., 2002). Activations within this ROI revealed by the contrasts of interest (see below) were thresholded at $p < 0.001$ (uncorrected) for display purposes and only local maxima surviving an FDR-corrected threshold of $p < 0.01$ within the ROI are reported.

The design matrix for each participant consisted of 18 regressors. All regressors collapsed across the direction of the coherent ramps; i.e., 100% coherent segments in which the ramps moved up were collapsed with 100% coherent segments in which the ramps moved down. The first regressor modeled the hemodynamic response to the onset of each sound block as a stick function (i.e., delta function with 0 s duration). Regressors 2–7 modeled the onset and duration of the segments within a block corresponding to one of the six levels of spectrotemporal coherence (0, 20, 40, 60, 80, and 100%). Regressors 8–18 modeled the response to changes in coherence as stick functions, with the eighth regressor modeling 0% changes (i.e., all consecutive percentage coherence pairs of 0-0,

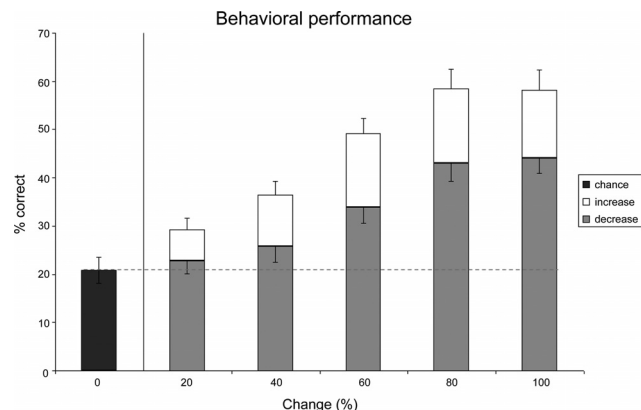


Figure 2. Behavioral results. Mean behavioral performance (\pm SEM) for detecting changes in coherence in the magnetic resonance imaging scanner. The black bar represents false alarm button presses for contiguous segments without a change in coherence ("0% change"), white bars indicate performance for changes across which coherence increased, and gray bars indicate performance for changes across which coherence decreased.

20-20, 40-40, 60-60, 80-80, 100-100), whereas the subsequent regressors modeled "positive" and "negative" changes of a given magnitude (+20, -20, +40, -40, +60, -60, +80, -80, +100, -100%). The 6 s silent baseline epochs between sound blocks were not modeled explicitly in the design matrix.

The following planned contrasts were performed. To delineate the response to sound onset, a simple contrast, [1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0], was performed. This contrast was used to derive an ROI of bilateral auditory cortex and is orthogonal to the contrasts of interest. The bilateral ROI had a size of 10,124 voxels (4934 voxels in left auditory cortex, 5190 voxels in right auditory cortex). To probe for an effect of increase in activity with increasing absolute coherence, the corresponding regressors 2–7 (i.e., 0 to 100% coherence) were linearly weighted as follows: [0 -2.5 -1.5 -0.5 0.5 1.5 2.5 0 0 0 0 0 0 0 0 0 0].

To probe for an effect of increasing change in coherence, the corresponding regressors 8–18 (i.e., 0, +20, -20...+100, -100% change magnitude) were linearly weighted [0 0 0 0 0 0 -2.73 -1.73 -1.73 -0.73 -0.73 0.27 0.27 1.27 1.27 2.27 2.27]. These weighted values are all mean centered on zero.

The behavioral performance (Fig. 2) for detecting the various types of changes (i.e., 0, +20, -20...+100%, -100%) was as follows: [20.82 29.18 22.83 36.42 25.73 49.22 33.81 58.40 43.03 58.23 44.12].

To probe for an effect of relative texture salience that reflects the behavioral asymmetry in which positive changes are more readily detected than negative changes (Fig. 2), we computed the difference between the detection rates for positive and negative changes. The differences between detecting +20 and -20%, +40 and -40%, +60 and -60%, +80 and -80%, and +100 and -100% changes were [6.35, 10.7, 15.42, 15.37, 14.12], respectively; thus, the corresponding contrast probing for an effect of relative texture salience weighed the five different positive and negative change pairs [0 0 0 0 0 0 0 6.35 -6.35 10.7 -10.7 15.42 -15.42 15.37 -15.37 14.12 -14.12].

This contrast is also mean centered on zero and reflects the magnitude of the perceptual difference between positive and negative changes.

Results

Behavioral results obtained during scanning for the detection of a change in spectrotemporal coherence between textures are shown in Figure 2. Performance increased with the magnitude of change (both for boundaries across which coherence increased or decreased) and was significantly better than chance or false alarm performance corresponding to 0% changes; two separate repeated-measures ANOVAs with factor ChangeLevel (0, 20, 40, 60, 80, 100%) for changes across which coherence either increased or decreased revealed main effects of ChangeLevel (in-

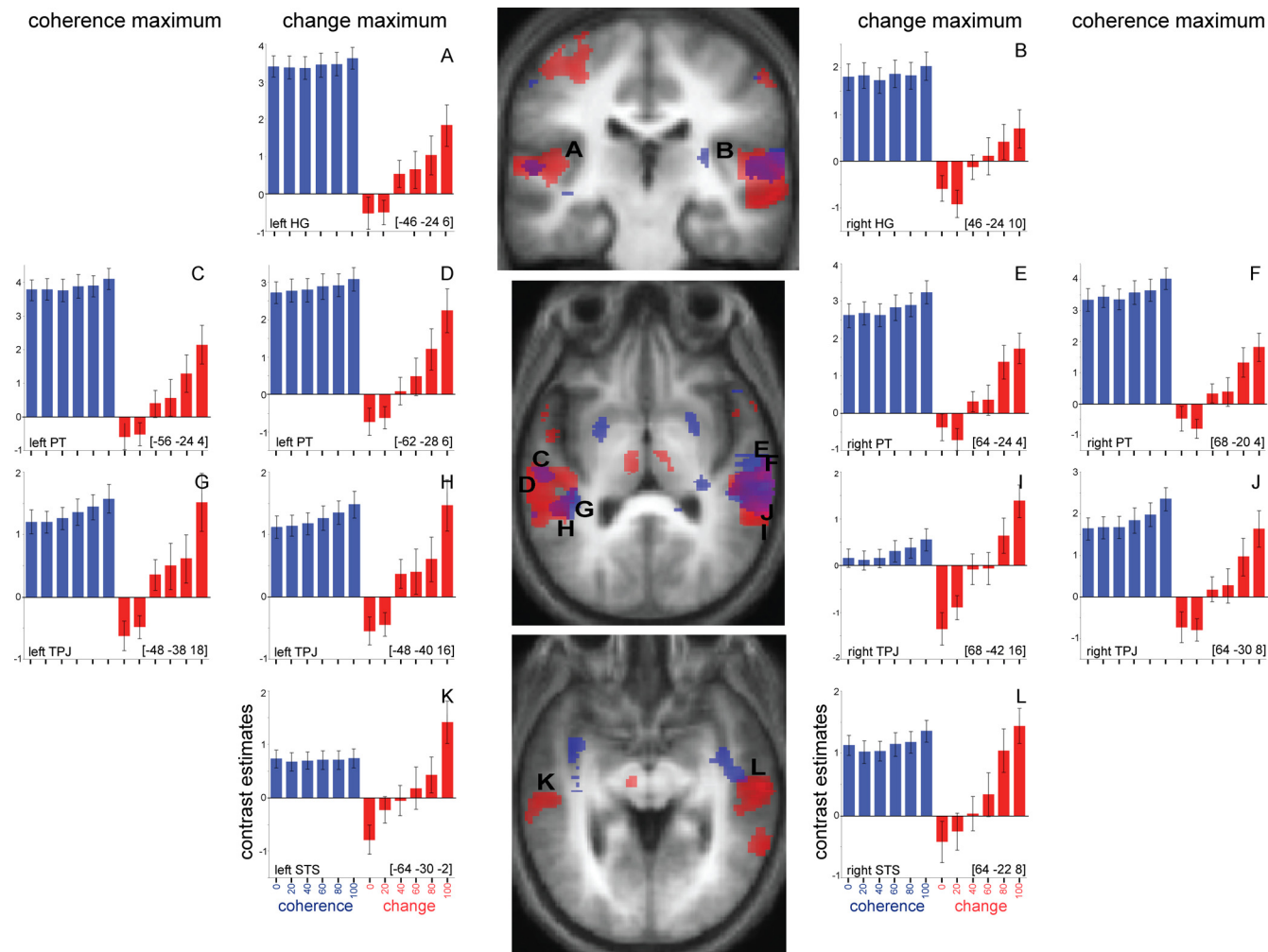


Figure 3. Areas showing an increased hemodynamic response as a function of increasing absolute coherence (blue) and increasing change in coherence (red). Results are rendered on coronal ($y = -24$; top) and tilted [pitch = -0.5 ; middle (superior temporal plane) and bottom (STS)] sections of participants' normalized average structural scans. The bar charts show the mean contrast estimates (\pm SEM) in a sphere with 10 mm radius around the local maximum corresponding to the six levels of absolute coherence (blue) and the six levels of change in coherence (red). Change in coherence levels are pooled across positive and negative changes so as to show the main effect of change in coherence magnitude. The charts nearest the brain show the mean response in the sphere around the local maxima for increasing change in coherence; those at the sides show the mean response in the sphere around local maxima for increasing absolute coherence. Note that the placement of the identifying letter (A–L) in the brain sections only approximate the precise stereotactic $[xyz]$ coordinates at the bottom corner of each chart (A–L, respectively), because no single planar section can contain all the local maxima simultaneously.

crease) [$F_{(5,110)} = 58.0, p < 0.001$] and ChangeLevel (decrease) [$F_{(5,110)} = 23.04, p < 0.001$], respectively. Pairwise comparisons (two-tailed t tests) with performance for 0% changes were all significant ($p < 0.05$) for change levels from 20 to 100% (increase) and 60–100% (decrease), indicating that participants performed above chance for these changes in spectrotemporal coherence. Furthermore, performance was better for changes across which coherence increased; a repeated-measures ANOVA with factors ChangeLevel (0–100%) and ChangeType (increase vs decrease) revealed main effects for ChangeLevel ($F_{(5,110)} = 52.05, p < 0.001$) and ChangeType [$F_{(1,22)} = 52.32, p < 0.001$], as well as a significant interaction [$F_{(5,110)} = 7.87, p < 0.001$].

We carried out an analysis to identify areas in auditory cortex that parametrically varied in activity as a function of the absolute magnitude of change in coherence (i.e., for both positive and negative changes in coherence) at the boundaries between adjacent segments. The analysis revealed BOLD signal increases in Heschl's gyrus (HG), planum temporale (PT), temporoparietal junction (TPJ), and superior temporal sulcus (STS) as a function of absolute change magnitude (Fig. 3, Table 1). The bar charts in

Figure 3 show (in red) the contrast estimates for the different degrees of change in coherence at the boundaries between textures in all these areas of auditory cortex.

Next, we sought areas within auditory cortex that were increasingly activated as a function of increasing spectrotemporal coherence within textures. Bilateral areas in auditory association cortex, including PT and its extension into TPJ, showed a contrast estimate increase with increasing absolute spectrotemporal coherence (Fig. 3, blue bars; Table 1). In contrast, activity in HG and STS did not differ significantly across the six levels of spectrotemporal coherence. To determine whether the absence of an effect for increasing spectrotemporal coherence in HG and STS was not an artifact of the statistical threshold used and was qualitatively different from the response in those areas to the magnitude of change, we performed repeated-measures ANOVA based on the mean contrast estimate in a sphere with 10 mm radius around the local maxima in HG and STS corresponding to the six levels of change in coherence. Two separate 2 Hemisphere (left vs right) \times 2 Condition (coherence vs change) \times 6 Level (1–6) repeated measures ANOVAs for HG and STS revealed significant

Table 1. Stereotactic MNI-coordinates

Contrast	Area	Left hemisphere				Right hemisphere			
		x	y	z	t value	x	y	z	t value
Change in coherence	HG	−46	−24	6	4.86	46	−24	10	3.8
	PT	−62	−28	6	5.45	64	−24	4	6.94
	TPJ	−48	−40	16	6.34	68	−42	16	7.63
		66	−44	14	7.68				
	STS	−64	−30	−2	4.55	64	−22	−8	6.16
Absolute coherence	50				2			8.07	
	PT	−56	−24	4	4.58	68	−20	4	6.71
Increasing versus decreasing coherence change	TPJ	−48	−38	18	4.91	64	−30	8	7.39
	TPJ	−46	−32	20	10.36	66	−28	20	7.10
−54		−34	22	9.78	46	−26	14	6.20	
−42		−20	16	8.77					

Local maxima for the effects of increasing change in coherence, increasing absolute coherence, and increasing versus decreasing coherence change are shown. Results are thresholded at $p < 0.01$, FDR corrected.

Condition \times Level interactions: $F_{(5,110)} = 4.54$, $p = 0.001$ for HG, and $F_{(5,110)} = 7.64$, $p < 0.001$ for STS, respectively.

The experimental design also enabled us to investigate in more detail an effect of absolute spectrotemporal coherence by way of changes in relative coherence. Behavioral results (Fig. 2) demonstrated that boundaries across which coherence increased (positive changes) were generally more salient than those across which coherence decreased (negative changes), suggesting a perceptual asymmetry in which texture salience increases with spectrotemporal coherence. We tested whether this perceptual asymmetry (Cusack and Carlyon, 2003) was also reflected at the neural level (see Materials and Methods). Figure 4 shows that this was the case in TPJ, which showed stronger responses to increases as opposed to decreases in spectrotemporal coherence between textures (see also Table 1).

Discussion

The results demonstrate a specific mapping of abstracted object properties, as represented by spectrotemporal coherence, and object boundaries, as represented by changes in spectrotemporal coherence, to distinct regions of auditory cortex. First, activity in auditory cortex, including HG, PT, TPJ, and STS, increased as a function of the magnitude of the change in spectrotemporal coherence at boundaries between textures. Second, activity as a function of the absolute spectrotemporal coherence within textures increased in auditory association cortex in PT and in TPJ. Finally, increases in spectrotemporal coherence at segment boundaries were more perceptually salient than decreases in spectrotemporal coherence at object boundaries, and this was reflected by stronger neural activity at such boundaries.

Although the observed parametric responses to absolute spectrotemporal coherence within textures and change in coherence between textures show some overlap in cortical resources (in PT and TPJ), they are indeed separable processes, because the experimental design orthogonalized the absolute coherence of acoustic textures and changes in coherence at boundaries. This indicates that the overlapping representations of change in coherence and absolute coherence in the nonprimary auditory areas in PT and TPJ represent a distinct mapping of these two processes in similar cortical areas; these mappings could be subserved by activity within distinct units or networks in those areas (Price et al., 2005; Nelken, 2008; Nelken and Bar-Yosef, 2008). Furthermore, the

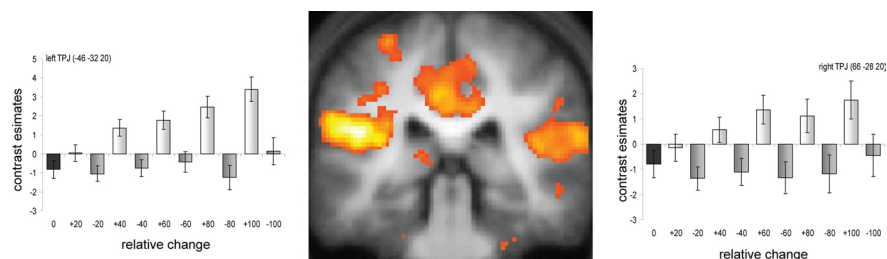


Figure 4. Increasing versus decreasing coherence change between texture segments. Coronal ($y = -32$) section showing areas that display a stronger increase for changes across which coherence increased than for changes across which coherence decreased. The bar charts at the sides show the mean contrast estimates (\pm SEM) in a sphere with 10 mm radius around the local maxima in left and right TPJ, respectively, for the different conditions.

results are unlikely to be confounded by the behavioral task (detection of a change in spectrotemporal coherence), because a pilot study in which the absolute spectrotemporal coherence of the sounds was task relevant yielded very similar results (supplemental Fig. S1, available at www.jneurosci.org as supplemental material).

A number of neuronal mechanisms might underlie the response to the boundaries that we demonstrate across auditory cortex (including primary cortex) and to the acoustic texture coherence that we demonstrate in association cortex. Computationally, both boundary detection and acoustic texture analysis within boundaries must depend on the statistical properties of the stimulus over frequency–time space, because low-level acoustic features such as spectral density over time were kept constant. For the present stimuli, boundary detection requires mechanisms that do not need to assess large spectrotemporal regions but still need to assess a “local” statistical rule change in the absence of any physical “edge” (as would be the case in the perception of an object arising out of silence, for example, where a boundary could be defined by a discontinuity in intensity). Acoustic texture coherence analysis necessarily involves larger spectrotemporal regions than boundary detection, and the analysis of boundary before texture that we demonstrate here is consistent with the idea that more extended segments of spectrotemporal space are analyzed in areas further from primary cortex. This notion is further supported by studies focusing on the time domain that suggest that the analysis of sound occurs over longer time windows in nonprimary cortex than in primary cortex (Boemio et al., 2005; Overath et al., 2008). In terms of the underlying neuronal mechanism for the present stimulus, we are not aware of any studies of coherent FM in either primary or nonprimary cortex. Neurons that are sensitive to the direction of single-FM sweeps

have been demonstrated in the auditory cortex of rats (Ricketts et al., 1998), cats (Mendelson and Cynader, 1985; Heil et al., 1992), and rhesus monkeys (Tian and Rauschecker, 2004) (for a review, see Rees and Malmierca, 2005). In our study, the analysis of boundaries (in primary cortex and association cortex) and texture (in association cortex) could be subserved by ensembles of such units tuned to similar sweeps in different regions of frequency–time space. Alternatively, if such neurons were ever shown to exist, boundaries and acoustic texture could also be analyzed by single neurons that were sensitive to coherent FM over spectrotemporal regions. In the case of both ensemble mechanisms and single-neuron mechanisms, the “receptive field” of the mechanism would need to be larger for texture analysis in association areas than for boundary detection in primary areas.

The present study provides a contrasting yet complementary approach to change detection mechanisms from the classical mismatch negativity paradigm, which is thought to reflect the violation of a previously established regularity (Näätänen and Winkler, 1999). Both paradigms require mismatch or change detection processes in auditory cortex. However, our results suggest that, in the current stimulus paradigm, the emergence of regularity (or coherence) has a different representation to its disappearance or violation. For example, the transition from noise to a regular interval sound with pitch has a different cortical representation than the reverse transition (Krumbholz et al., 2003). Recently, Chait et al. (2007, 2008) demonstrated distinct cortical mechanisms for the detection of auditory edges based on statistical properties, where the detection of a statistical regularity (in violation of a previous irregularity) had a different cortical signature than the detection of a violation of statistical regularity. The current results support the existence of such neural and perceptual asymmetries. We propose that the degree of spectrotemporal coherence is encoded in a continuous manner, with neurons tuned to coherence levels that are equal or greater in coherence than the thresholds of the neurons. Such a cumulative neural code contains an inherent asymmetry (Treisman and Gelade, 1980; Cusack and Carlyon, 2003); transitions to more coherent sounds excite a larger neural population, rendering them more perceptually salient. This is reflected in the neural response (Fig. 4).

The stimulus manipulation used in this study addresses generic processes underlying complex auditory object analysis, but it is not intended to represent all possible auditory object classes. In speech perception, the spectrotemporal analysis necessarily spans a large frequency range over multiple temporal scales, where coherent acoustic properties (or those with “common fate”) need to be abstracted across the frequency–time axis; for example, formant transitions often display such coherent acoustic spectrotemporal properties (Stevens, 1998). At the same time, there is no one ecological sound that the acoustic texture stimulus represents, but we argue here that its generic nature ensures applicability to a variety of ecological sound properties. Whereas coherent FM is arguably a relatively weak grouping cue compared to simultaneous onset and harmonicity (Carlyon, 1991; Summerfield and Culling, 1992; Darwin and Carlyon, 1995), it is nevertheless one basis upon which figure–ground selection can occur (McAdams, 1989). It is important to note, however, that these studies generally used simple sinusoidal FM in which “FM coherence” was defined as either in-phase or out-of-phase. The stimulus employed here is more complex in the sense that spectrotemporal coherence can only be detected as a whole, regardless of low-level features such as phase, because FM ramps were randomly distributed in frequency–time space.

The visual depiction of the stimulus (Fig. 1) evokes the coherent visual motion paradigm using random dot kinematograms (Newsome and Paré, 1988; Britten et al., 1992; Rees et al., 2000; Braddick et al., 2001). However, direct comparisons with the visual system based on superficial similarities are often not straightforward and need to be treated with caution (King and Nelken, 2009). For example, objects in the visual stimulus are defined spatially, whereas space plays a relatively minor role in the definition of auditory objects. Furthermore, in the present case, the perceptual effect is more subtle than in the visual domain.

The data reported here move beyond the analysis of simple FM sounds to the analysis of auditory object patterns within stochastic stimuli; such auditory object analysis is dependent on mechanisms that are fundamental for the analysis of ecologically valid sounds in a dynamic auditory environment. We demonstrate a mechanism for the assessment of acoustic texture boundaries that is already present in primary auditory cortex, based on recognizing changing, higher-order statistical properties governing frequency–time space. Such a mechanism precedes the encoding of the absolute properties of acoustic textures in higher-level auditory association cortex.

References

- Belin P, Zatorre RJ, Hoge R, Evans AC, Pike B (1999) Event-related fMRI of the auditory cortex. *Neuroimage* 10:417–429.
- Boemio A, Fromm S, Braun A, Poeppel D (2005) Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nat Neurosci* 8:389–395.
- Braddick OJ, O’Brien JMD, Wattam-Bell J, Atkinson J, Hartley T, Turner R (2001) Brain areas sensitive to coherent visual motion. *Perception* 30:61–72.
- Britten KH, Shadlen MN, Newsome WT, Movshon JA (1992) The analysis of visual motion: a comparison of neuronal and psychophysical performance. *J Neurosci* 12:4745–4765.
- Carlyon RP (1991) Discriminating between coherent and incoherent frequency modulation of complex tones. *J Acoust Soc Am* 89:329–340.
- Chait M, Poeppel D, de Cheveigné A, Simon JZ (2007) Processing asymmetry of transitions between order and disorder in human auditory cortex. *J Neurosci* 27:5207–5214.
- Chait M, Poeppel D, Simon JZ (2008) Auditory temporal edge detection in human auditory cortex. *Brain Res* 1213:78–90.
- Cusack R, Carlyon RP (2003) Perceptual asymmetries in audition. *J Exp Psychol Hum Percept Perform* 29:713–725.
- 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.
- Darwin CJ, Carlyon RP (1995) Auditory grouping. In: *Handbook of perception and cognition*, Vol 6, Hearing (Moore BCJ, ed), pp 387–424. London: Academic.
- 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.
- Genovese CR, Lazar NA, Nichols T (2002) Thresholding of statistical maps in functional imaging using the false discovery rate. *Neuroimage* 15:870–878.
- Griffiths TD, Warren JD (2004) What is an auditory object? *Nat Rev Neurosci* 5:887–892.
- 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.
- Heil P, Rajan R, Irvine DR (1992) Sensitivity of neurons in cat primary auditory cortex to tones and frequency-modulated stimuli. I: Effect of variation of stimulus parameters. *Hear Res* 63:108–134.

- 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.
- King AJ, Nelken I (2009) Unraveling the principles of auditory cortical processing: can we learn from the visual system? *Nat Neurosci* 12:698–701.
- Krumbholz K, Patterson RD, Seither-Preisler A, Lammertmann C, Lütkenhöner B (2003) Neuromagnetic evidence for a pitch processing center in Heschl's gyrus. *Cereb Cortex* 13:765–772.
- Kubovy M, Van Valkenburg D (2001) Auditory and visual objects. *Cognition* 80:97–126.
- McAdams S (1989) Segregation of concurrent sounds. I. Effects of frequency-modulation coherence. *J Acoust Soc Am* 86:2148–2159.
- Mendelson JR, Cynader MS (1985) Sensitivity of cat primary auditory-cortex (A1) neurons to the direction and rate of frequency modulation. *Brain Res* 327:331–335.
- Näätänen R, Winkler I (1999) The concept of auditory stimulus representation in cognitive neuroscience. *Psychol Bull* 125:826–859.
- Nelken I (2004) Processing of complex stimuli and natural scenes in the auditory cortex. *Curr Opin Neurobiol* 14:474–480.
- Nelken I (2008) Processing of complex sounds in the auditory system. *Curr Opin Neurobiol* 18:413–417.
- Nelken I, Bar-Yosef O (2008) Neurons and objects: the case of auditory cortex. *Front Neurosci* 2:107–113.
- Newsome WT, Paré EB (1988) A selective impairment of motion perception following lesions of the middle temporal visual area (MT). *J Neurosci* 8:2201–2211.
- Ohl FW, Scheich H, Freeman WJ (2001) Change in pattern of ongoing neural activity with auditory category learning. *Nature* 412:733–736.
- Overath T, Kumar S, von Kriegstein K, Griffiths TD (2008) Encoding of spectral correlation over time in auditory cortex. *J Neurosci* 28:13268–13273.
- Price C, Thierry G, Griffiths T (2005) Speech-specific auditory processing: where is it? *Trends Cogn Sci* 9:271–276.
- Rees A, Malmierca MS (2005) Processing of dynamic spectral properties of sounds. *Int Rev Neurobiol* 70:299–330.
- Rees G, Friston K, Koch C (2000) A direct quantitative relationship between the functional properties of human and macaque V5. *Nat Neurosci* 3:716–723.
- Ricketts C, Mendelson JR, Anand B, English R (1998) Responses to time-varying stimuli in rat auditory cortex. *Hear Res* 123:27–30.
- Scholte HS, Jolij J, Fahrenfort JJ, Lamme VA (2008) Feedforward and recurrent processing in scene segmentation: electroencephalography and functional magnetic resonance imaging. *J Cogn Neurosci* 20:1–13.
- Schönwiesner M, Novitski N, Pakarinen S, Carlson S, Tervaniemi M, Näätänen R (2007) Heschl's gyrus, posterior superior temporal gyrus, and mid-ventrolateral prefrontal cortex have different roles in the detection of changes. *J Neurophysiol* 97:2075–2083.
- Stevens KN (1998) *Acoustic phonetics*. Cambridge, MA: MIT.
- Summerfield AQ, Culling J (1992) Auditory segregation of competing voices: absence of effects of FM or AM coherence. *Philos Trans R Soc Lond B Biol Sci* 336:357–366.
- Tian B, Rauschecker JP (2004) Processing of frequency-modulated sounds in the lateral auditory belt cortex of the Rhesus monkey. *J Neurophysiol* 92:2993–3013.
- Treisman AM, Gelade G (1980) A feature-integration theory of attention. *Cognit Psychol* 12:97–136.
- Zatorre RJ, Bouffard M, Belin P (2004) Sensitivity to auditory object features in human temporal neocortex. *J Neurosci* 24:3637–3642.