nature neuroscience

Hierarchical and asymmetric temporal sensitivity in human auditory cortices

Anthony Boemio[1], Stephen Fromm[2], Allen Braun[2] & David Poeppel[3]

Lateralization of function in auditory cortex has remained a persistent puzzle. Previous studies using signals with differing spectrotemporal characteristics support a model in which the left hemisphere is more sensitive to temporal and the right more sensitive to spectral stimulus attributes. Here we use single-trial sparse-acquisition fMRI and a stimulus with parametrically varying segmental structure affecting primarily temporal properties. We show that both left and right auditory cortices are remarkably sensitive to temporal structure. Crucially, beyond bilateral sensitivity to timing information, we uncover two functionally significant interactions. First, local spectrotemporal signal structure is differentially processed in the superior temporal gyrus. Second, lateralized responses emerge in the higher-order superior temporal sulcus, where more slowly modulated signals preferentially drive the right hemisphere. The data support a model in which sounds are analyzed on two distinct timescales, 25–50 ms and 200–300 ms.

Structure, function and lateralization in human auditory cortex are the focus of much recent work[1]. One central issue concerns the origin and nature of lateralization. For example, there are subtle anatomic and physiological asymmetries in the afferent pathway, but compelling functional asymmetries attributable to cortical processing[2,3]. Where do such asymmetries originate? One hypothesis proposes that functional lateralization arises from differences in the early spectrotemporal computations performed in auditory cortices that transform sensory representations of signals into more abstract perceptual codes. A prevailing model is that temporal features are processed predominantly in the left hemisphere and spectral features in the right[4]. A second and different source of lateralization derives from the nature of the stored representations that the transformed sensory information must interface with for further processing—for example, lexical information in the left and affective prosodic information in the right hemisphere.

Both explanations have led to the notion that speech—whether resulting from lateralization of stored lexical representations or from early auditory cortical specialization for processing temporal signal attributes—is preferentially processed within the left hemisphere[5–7], whereas processing of dynamic pitch and prosody—whether resulting from lateralized representation of higher-order phrase-level intonation or specialized analysis of spectral information—is carried out in the right hemisphere[8–10].

We argue that both hemispheres together—including left and right non-primary auditory areas—participate in one critical intermediate computation, the analysis of the auditory signal on multiple timescales[3,11,12], with the relevant scales being 25–50 ms and 200–300 ms (ref. 3). In addition, we propose that functional lateralization emerges from differential connectivity patterns linking temporal cortices along the afferent pathway such that information processed on the longer timescales is routed predominantly to higher-order right hemisphere cortices, whereas information resulting from processing on the shorter timescale primarily projects to the left. To evaluate these hypotheses, we varied a single stimulus para-meter, the temporal structure, and looked for differential activation along the afferent pathway and between the two hemispheres. The present design controls for potential spectral confounds in a fashion that was not possible in previous studies in which stimuli varied both temporally and spectrally[4,13].

Fifteen participants listened passively to non-speech stimuli (**Fig. 1**) while we recorded the hemodynamic responses from the entire brain using a single-trial sparse acquisition fMRI design[14,15]. We created 9-s auditory signals by concatenating short-duration narrowband noise segments, spanning a range of segment transition rates from 3 to 83 segments per second, encompassing the syllabic to segmental transition rates of speech[16] (**Fig. 1a**,**b**; examples can be heard in **Supplementary Audios 1–5** online). Each segment had a bandwidth of 125 Hz and a segment center frequency spanning a half-octave range from 1,000 Hz to 1,500 Hz. The bandwidth was chosen to be within the critical band at that frequency[17] and interpretable in the context of the rate and bandwidth of speech formants[16]. Local spectrotemporal variations were introduced by constructing two types of segments. In one, the frequency remained constant throughout the signal

[1]Laboratory of Brain and Cognition, National Institute of Mental Health, National Institutes of Health, Bethesda, Maryland 20892, USA. [2]National Institute of Deafness and Other Communication Disorders, National Institutes of Health, Bethesda, Maryland 20892, USA. [3]Department of Linguistics and Department of Biology, University of Maryland, College Park, Maryland 20742, USA. Correspondence should be addressed to D.P. (dpoeppel@deans.umd.edu).

Published online 20 February 2005; doi:10.1038/nn1409

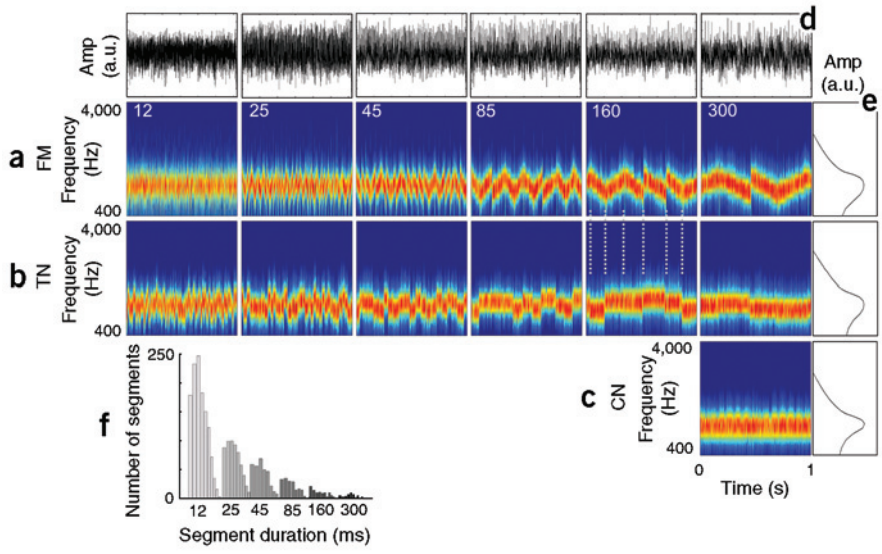


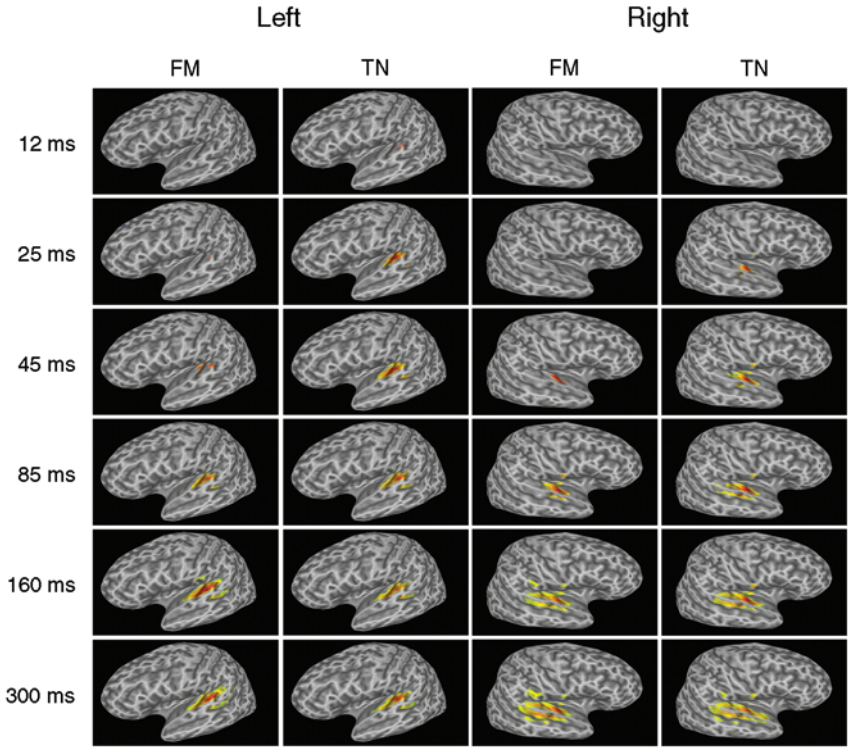
Figure 1 Concatenated narrow-band noise stimuli. (**a**,**b**) Spectrograms of (**a**) 6 FM type stimuli and (**b**) 6 TN type stimuli. Frequency (Hz) is plotted on the ordinate, time (ms) on the abscissa. Mean segment duration (ms) is shown at the top of the FM plots. Dotted lines connecting the segments in the 160 ms plots denote the common segmental structure of FM and TN type stimuli. (**c**) Spectrogram of CN stimulus. Although all 13 conditions were 9 s long in the study, only 1 s is shown here for clarity. (**d**) Temporal profiles of the six FM stimuli, obtained by summing over frequency, showing the common amplitude envelope across segment duration. (**e**) Spectral profiles of the 300-ms FM and TN segment types and CN stimulus, obtained by collapsing over time, showing the similar long-term spectrum across segment type. Amplitude (Amp) values are in arbitrary units (a.u.). (**f**) Number of segments contained in each of the six FM and TN conditions as a function of segment duration. Distributions are Gaussian with clipped tails to avoid overlap between conditions. Note the log scale on the abscissa.

(TN; **Fig. 1b**); in the other, frequency was swept linearly upward or downward randomly (FM; **Fig. 1a**). A control stimulus (CN; **Fig. 1c**) was constructed from a single 9-s TN segment with center frequency in the middle of the half-octave range.

We show that early and higher-order auditory cortical areas are exquisitely sensitive to temporal structure bilaterally. In addition, local spectro-temporal structure is differentially processed within the superior temporal gyrus. Finally, in higher-order superior temporal sulcus, slowly modulated signals preferentially drive the right hemisphere. To account for these observations, we present a model involving cortical processing of auditory signals on short and long timescales, and a hypothesized differential connectivity pattern from lower- to higher-order auditory areas.

RESULTS

The stimuli were effective at activating auditory cortex selectively and robustly (**Fig. 2**). Two independent analyses of the fMRI data were carried out, one at the cohort level, the second a region-of-interest (ROI) analysis at the level of individual subjects. Both analyses were based on categorical contrasts between each of the six FM and TN stimuli and the one CN stimulus. The cohort analysis produced a single set of activation maps across all subjects, whereas the single-subject ROI analysis produced one set of maps for each subject. These contrasts were designed to identify cortical regions sensitive to segmental structure. Given the range of controls designed into the stimuli, segmental structure should be the primary feature driving the response.



Cohort analysis

Contrasts between the active FM and TN stimuli and the CN control yielded bilateral activations (SPM 99, $P < 0.05$, corrected) in the transverse temporal gyrus (TTG), superior temporal gyrus (STG) and superior temporal sulcus (STS). All areas showed a strong effect of segment duration, with longer durations yielding greater activation (**Fig. 2**). **Table 1** shows the Montreal Neurologic Institute (MNI) coordinates and number of suprathreshold voxels for all conditions producing suprathreshold activation. Several effects are visible: (i) the increasing spatial extent of activation from 12 to 300 ms for all conditions (ii) the greater activation observed for TN as compared to FM type for short segment duration, and (iii) the hemispheric asymmetry, in which

Figure 2 Surface-mapped activation from the cohort analysis shown on the inflated N27 brain. Segment duration (ms) is specified on the left, segment type (FM versus TN) and hemisphere at the top. Threshold corresponds to $P < 0.05$, corrected.

longer segment duration stimuli elicited greater activation in the right hemisphere.

Inspection of the activation maps show that the majority of the activation was limited to STG and STS, with the remainder being generated by TTG, the putative core human auditory cortex[18]. The small activation of TTG was not due to its smaller size relative to STG or STS, but to a paucity of suprathreshold voxels compared to the total number of voxels comprising this area. The modest activation is likely to be due to the nature of the contrast used, which was designed to identify areas that explicitly code temporal structure, because all stimuli had similar spectral envelope (**Fig. 1e**), temporal envelope (**Fig. 1d**) and RMS (root mean square) power. Thus, it seems that TTG codes signal properties common to all the stimuli, whereas areas STG and STS represent signal properties unique to the TN and FM types.

To obtain the perceptual correlate of the hemodynamic responses, subjects rated the perceptibility of the individual segments in an offline, single-interval, two-alternative forced-choice task. **Figure 3** shows the results displayed as the proportion of responses for which individual segments were perceived. These data are plotted with the normalized hemodynamic response, expressed as the total number of suprathreshold voxels from the cohort analysis. The two curves—which can be interpreted as transfer functions between the acoustic properties of the stimulus and the perceptual and physiological (hemodynamic) responses they elicit—are very similar, indicating a tight correlation between perception and underlying physiology as assessed here.

Single-subject ROI analysis

Responses in TTG, STG and STS were subsequently analyzed on a subject-wise basis in the three ROIs identified in the cohort analysis. ROIs were defined by Talairach Daemon classification of the union of suprathreshold voxels across all conditions from the cohort analysis. The mean contrast value (mean voxel strength) in all three ROIs was computed for each condition and subject and then submitted to a repeated-measures ANOVA with the factors hemisphere, segment type and segment duration. The results showed several robust effects. First, all three regions showed a compelling main effect of segment duration (TTG, $P < 0.001$; STG, $P < 0.001$; STS, $P < 0.001$; **Fig. 4a**), consistent with the cohort analysis. Second, a strong main effect of segment type was observed in STG ($P < 0.001$; **Fig. 4b**) but not in TTG ($P < 0.12$) or STS ($P < 0.13$), as was an interaction between segment type and segment duration ($P < 0.05$; **Fig. 4c**). The effect is greatest at 45 ms SOA (stimulus onset asynchrony) and falls off on both sides (**Fig. 4**, inset). Post-hoc testing showed that neighboring conditions rose to significance ($P < 0.05$) (asterisks) when compared to the 300 ms SOA condition. Third, and crucially, for STS alone, we observed a significant main effect of hemisphere ($P < 0.001$, **Fig. 4d**) as well as a significant interaction between segment duration and hemisphere ($P < 0.001$, **Figs. 4e** and **5**), with slowly varying signals preferentially driving the right hemisphere. In contrast, in TTG and STG, neither the main effect of hemisphere (TTG, $P < 0.86$; STG, $P < 0.33$) nor the duration × hemisphere interaction reached significance (TTG, $P < 0.14$; STG, $P < 0.07$).

In summary, the ROI analysis showed a progressive elaboration of the representation of the acoustic signal along the afferent pathway from TTG to STG to STS (**Table 2**), in which TTG was weakly activated, STG showed a sensitivity to segment type (or stimulus identity, based on local spectrotemporal cues), and STS showed a hemispheric asymmetry to segment duration (or stimulus rate, based on durational cues).

DISCUSSION

We focus on three important findings in turn: (i) the bilateral response to temporal structure (**Figs. 2** and **4a**) (ii) the differential sensitivity in STG to FM and TN signals (**Figs. 4b**,**c**) and (iii) the hemispheric asymmetry in STS as a function of segment duration (**Figs. 4e** and **5**).

Table 1 Cohort analysis summary

	Left		Right	
Seg dur (ms)	FM	TN	FM	TN
12	–	–	–	–
25	–	–56 –28 4 6.88 (64)	–	52 –16 0 6.36 (10)
45	–40 –36 8 4.91 (7)	–56 –24 4 8.16 (97)	48 –20 0 5.30 (10)	52 –12 0 8.95 (61)
85	–60 –28 8 7.14 (58)	–60 –28 4 8.36 (136)	52 –8 0 8.27 (44)	56 –12 0 8.8 (163)
160	–56 –24 4 10.94 (242)	–56 –28 4 8.61 (226)	52 –12 0 9.58 (256)	52 –12 0 9.77 (328)
300	–60 –28 8 10.60 (252)	–56 –28 10.38 (248)	452 –12 0 9.52 (314)	52 –12 0 8.56 (229)

Seg dur, segment duration. Activation is organized by condition and hemisphere. Values in the cells in the left-most column denote segment duration in ms. Remaining cells contain (i) the MNI coordinates of the most activated voxel for the specified condition (ii) the *t*-score of the most activated voxel, and (iii) the total number of suprathreshold voxels at $P < 0.05$ corrected across all areas (parentheses). No suprathreshold activation was observed at 12 ms SOA for any condition.

Bilateral sensitivity to temporal structure

A robust main effect of segment duration (**Fig. 4a**) was obtained in both hemispheres of all three ROIs, similar to the overall response pattern shown in the cohort analysis (**Fig. 2**). The observation that varying a narrowband signal along a single (temporal) dimension induces differential processing in non-primary auditory cortices substantially extends the range of cortex to which we must attribute temporal sensitivity. The finding is consistent with prior data[11,19] but new in demonstrating the extensive contribution of non-primary areas. Typically, core auditory fields[18] are studied to investigate the representation of signals[19–22]. We show that temporal lobe structures within and beyond belt and parabelt projection areas also reflect the temporal properties of the stimulus[19]. Furthermore, both left and right non-primary areas showed a high sensitivity to this temporal structure, suggesting that the rightward lateralization of spectral (for example, pitch change) sensitivity[2,4,8,9]

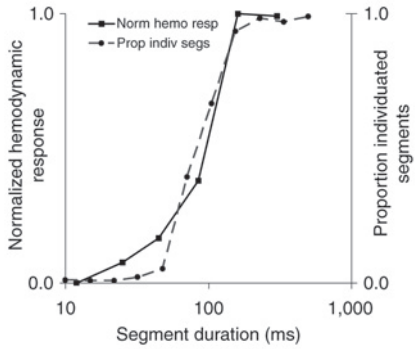


Figure 3 Comparison of physiological (cohort analysis) and behavioral responses. Normalized aggregate hemodynamic response (left, solid line) and the proportion of individuated segments judgments (right, dashed line) are plotted together to show the tight correlation between perception and neural representation of stimulus segmental structure. The hemodynamic response was computed by normalizing the number of suprathreshold voxels from the cohort analysis across all 12 contrasts. The proportion of individuated segments was obtained by summing the number of times that segments were classified as perceptible for each condition and for each subject, normalizing across conditions, and averaging across all subjects.

Table 2 Summary of single-subject ROI analysis

	Left				Right			
	Seg dur	Seg type	Hemi asym	Timescales	Seg dur	Seg type	Hemi asym	Timescales
TTG	Yes	No	No	<25 ms?	Yes	No	No	<25 ms?
STG	Yes	Yes	No	25–50 ms, 200–300 ms	Yes	Yes	No	25–50 ms, 200–300 ms
STS	Yes	No	Yes	≥200–300 ms	Yes	No	Yes	≥200–300 ms

Seg dur, segment duration; seg type, segment type; hemi asym, hemispheric asymmetry. All three areas (TTG, STG and STS) showed a robust effect of segment duration. STG alone showed a main effect of segment type and an interaction between segment type and segment duration. In STS, a main effect of hemisphere and an interaction of hemisphere and segment duration were observed. Time-scale columns denote the hypothesized period over which auditory information is combined across time and across frequency in cortex. Longer processing times are associated with greater displacement along the afferent auditory pathway (see Discussion for details).

does not come at the cost of right auditory areas' sensitivity to timing information—a finding that may have important implications for theories of speech perception.

STG sensitivity to local spectrotemporal features

STG was strongly activated in both the cohort (**Table 1**) and ROI analyses, where main effects of segment duration (**Fig. 4a**) and segment type (**Fig. 4b**) were observed, in addition to an interaction between segment type and segment duration (**Fig. 4c**). Collectively, these results show that left and right STG are sensitive to both the local temporal structure (segment duration) and the spectral structure (segment type) of the stimuli.

We argued that auditory signals are analyzed simultaneously on at least two independent timescales, namely 25–50 ms and 200–300 ms; here we provide the physiological basis for these claims. The notion of processing on a particular timescale here refers to the integration of auditory information over time and across frequency in cortex.

Evidence for processing on the 25–50 ms timescale derives primarily from the difference observed in the hemodynamic response as a function of both segment duration and segment type (**Fig. 4b**,**c**). This difference is greatest at 45 ms SOA (**Fig. 4c**, inset). The concept of a temporal window 25–50 ms in width—independently supported by psychophysical[3] and neurophysiological[21] considerations—provides an explanatory basis for this effect when considered in the context of the segmental structure of the stimuli. For long segment duration, an FM segment sweeps across the half-octave (1,000–1,500 Hz) frequency range in the longest amount of time, making its spectral slope (Hz/s) the least steep, and thus most like the TN segment, which has a slope of zero. At this segment duration, the smallest difference between the hemodynamic responses of the two segment types is observed (**Fig. 4c**). As the segment duration decreases, the difference between the two segment types within the integration window increases: the FM segment sweeps across the same frequency span in less time. At 45 ms SOA, the duration of our hypothetical window, exactly one FM segment 'fits' within the window (spanning the entire half-octave frequency range), making it maximally different from the TN segment, and this corresponds to the largest observed difference in hemodynamic response between the two segment types. For further decreases in segment duration, the window will contain more than one segment, and these begin to fuse into a single homogenous stream, reducing the difference between the neural representation of the two segment types and thus the differential hemodynamic response.

Although analysis on the 25–50 ms timescale is manifested as a difference between the hemodynamic responses of the FM and TN segments—suggesting processing dedicated to enhancing differences in transient signals like speech sounds—it might also reflect extraction of segment boundaries. Detection and explicit representation of the abrupt changes in spectrotemporal structure are important because these are potent cues in the estimation of the onset and offset of sound sources[23].

Evidence supporting processing on the timescale of 200–300 ms is derived from the asymptote in the hemodynamic response for SOA >160 ms, evident in both the cohort (**Fig. 3** and **Table 1**) and ROI (**Fig. 4c**) analyses, each of which were based on different measures of the hemodynamic response. Asymptotic behavior was observed in both the spatial extent (number of suprathreshold voxels) and magnitude (mean voxel strength) of activation. Here again the concept of a temporal analysis window—in which a restricted portion of the auditory representation is processed—provides the basis for interpreting the physiological response. However, the specific features computed on this timescale are likely to differ from those computed on the 25–50 ms scale. For example, analysis within a 200–300 ms window may be optimized to represent

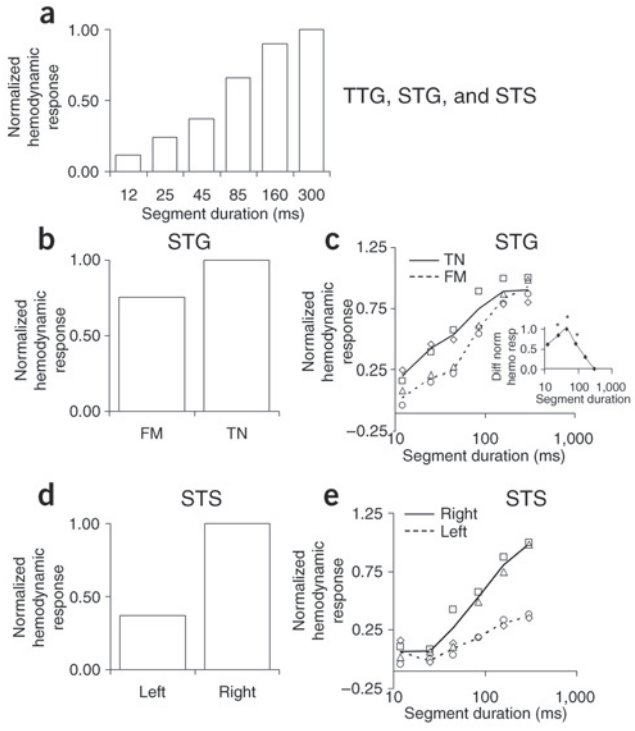


Figure 4 ROI analysis: STG sensitivity to segment type and STS hemispheric asymmetry. Normalized hemodynamic response (ordinate) for all plots represents the mean voxel strength of the contrast within the specified area. Values were averaged across subjects within each ROI and then normalized across conditions. When segment duration is plotted on the abscissa, it is in log scale. (**a**) Main effect of segment duration in all areas; longer segment duration produced greater activation. (**b**) Main effect of segment type in STG; overall, TN conditions produced greater activation than FM conditions. (**c**) Interaction between segment duration and segment type in STG (◇ left-TN, □ right-TN, ○ left-FM, Δ right-FM). Inset shows the difference between TN and FM conditions as a function of segment duration. The greatest difference occurs at 45 ms SOA. Asterisks denote significant differences relative to 300 ms SOA ($P < 0.05$). (**d**) Main effect of hemisphere in STS; overall, greater activation was observed in the right hemisphere. (**e**) Interaction between hemisphere and segment duration in STS (◇ left-TN, □ right-TN, ○ left-FM, Δ right-FM). The greatest disparity occurs at the longest segment durations where activation in right STS exceeds that in the left (**Fig. 5**).

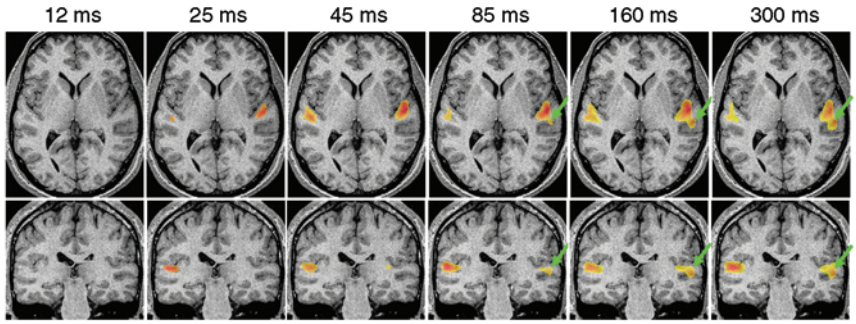


Figure 5 STS activation from cohort analysis. Axial and coronal slices for all six TN stimuli shown in neurological convention. Slices were chosen by finding the maximum activation in right STS for the 300 ms condition (MNI coordinates 61, –10, –16). Comparison between the left and right hemisphere in the axial slices shows that two distinct clusters of activation exist only in the right hemisphere; the more anterior in STG, the more posterior in STS (arrows). Coronal slices show that STS activation is confined to the dorsal bank (arrows) and occurs only for the three longest segment durations. Similar patterns are observed for the FM stimuli (data not shown). Activation maps for individual subjects derived from the ROI analysis (**Supplementary Fig. 1** online) show that the hemispheric asymmetry observed in STS in the cohort analysis is also observed at the individual subject level.

perceptual objects such as syllables and to mediate the processing of dynamic pitch and sentential prosody. In contrast, processing on the 25–50 ms timescale may be involved in extracting segment boundaries subserving (sub)segmental processing. Thus, the existence of two independent processing timescales facilitates the extraction of different features in a manner not possible with a single timescale. This greatly increases the efficacy of the transformation from an early sensory code—isomorphic with the acoustic signal—to a more stable perceptual code—isomorphic with the percept—comprising features extracted during earlier processing.

The longest segment duration used in this study was 300 ms, and thus it is not possible to determine with certainty what happens for segment durations >300 ms. Will the asymptote in hemodynamic response and number of suprathreshold voxels observed between 160–300 ms persist, or will these values ultimately decrease to baseline? To answer this question, recall that the contrasts between the 12 active FM and TN conditions and the 1 CN control condition compared the hemodynamic response elicited by stimuli with many segments to a stimulus with just one (9-s) segment. Thus, it is certain that, if the segment duration of the FM and TN conditions were increased to values approaching 9 s, the contrast between these and the CN condition eventually would fall to zero because the hypothetical contrast would be comparing increasingly similar entities. So the question becomes, "Is the observed asymptote in response a relative maximum or an absolute maximum?" Based on previous electrophysiological[24–26] and psychophysical[27] evidence, we predict that the asymptote between 160–300 ms is an absolute maximum and that the hemodynamic response will decrease to baseline as the segment duration is increased beyond 300 ms.

Hemispheric asymmetry in STS

STS—recently described as constituting the fourth level of processing in the primate auditory system[28]—showed a strong main effect of segment duration, like STG, demonstrating sensitivity to temporal structure over a range of more than an order of magnitude. In contrast to STG, however, STS was found not to be sensitive to segment type, and it showed a smaller asymptote in hemodynamic response for segment duration >160 ms. Yet a marked hemispheric asymmetry was observed (**Figs. 4d** and **5**), as was an interaction between hemisphere and segment duration (**Figs. 4e** and **5**) such that stimuli with segment duration >85 ms SOA produced robust activation in the dorsal bank of right STS, in both the cohort and single-subject analyses (**Supplementary Fig. 1** online). This asymmetry cannot be accounted for by appealing to signal complexity—because the stimuli used varied only along the temporal axis—or to speech specificity—given that the stimuli were neither speech signals nor perceived as such. In a recent study addressing a similar issue[4], a lateralization was observed consistent with that observed here. In that study, however, both temporal and spectral signal attributes were varied; we demonstrate that varying just a single dimension, temporal structure, can drive lateralization.

Hemispheric asymmetries of the auditory cortex have been documented in anatomical[29,30] and physiological studies, and it has been argued that they underlie lateralized auditory functions including the analysis of speech sounds[31–34] and pitch[8,9,35]. The right-hemisphere advantage for slow modulation rate (long segment duration) observed in STS in the present study—although similar to findings from research on phrase-level prosody[36] and the analysis of dynamic pitch—suggests a more general explanation.

The absence of sensitivity to segment type, the strong dependence on segment duration, and the slower progression to asymptote in the hemodynamic response between 160–300 ms relative to STG suggest that the integration of information in STS in both hemispheres occurs on a timescale equal to or greater than that in STG, that is ≥200–300 ms. To explain the observed hemispheric asymmetry, we hypothesize that left and right STS receive input differentially from STG through intra- and inter-hemispheric (transcallosal) fibers. Specifically, left STS receives contributions from left and right STG weighted toward processing on the 25–50 ms timescale, and right STS receives input weighted toward processing on the 200–300 ms timescale (**Supplementary Fig. 2** online). Evidence supporting this claim is reflected in the markedly different slopes of the mean voxel strength in left and right STS as a function of segment duration (**Fig. 4e**). Two factors are likely to contribute to this disparity: the difference in the magnitude of the input received from the two populations in STG, and the differential effect of the two temporal windows. For long segment duration, the greater activation in right STS is a consequence of the larger magnitude of the predominantly 200–300 ms input received from STG. This results from integration over a longer period relative to that received by the less activated left STS, which is smaller owing to integration in STG over only 25–50 ms. However, as the segment duration becomes shorter, the magnitude of the output from the 200–300 ms population in STG falls off faster than that from the 25–50 ms population, because individual short segments begin to fuse sooner in the longer window.

The appeal of a model based on differential cortical connectivity between areas STG and STS in explaining the asymmetry observed is fourfold. First, it is automatic in that the spectrotemporal structure of the stimulus drives the hemispheric asymmetry by selectively engaging two intrinsic processing timescales. Second, the confluence in STS of the independent representations of a signal, each extracted on different

timescales, to a single timescale in which the features from each representation are in temporal register facilitates the construction of more abstract representations of the signal. Third, our model explains why no obvious anatomic or physiologic asymmetries have been found in higher-order cortices—none are implicated. The model stipulates only the pattern of intra- and inter-hemispheric connectivity. No constraints are placed either on the spatial distribution of circuits within STG that mediate processing on each timescale or on the fine structure of the afferents that route the resulting information to STS. Fourth, although an architecture based on differential cortical connectivity differs mechanistically from previous proposals, it is consistent with them both conceptually and empirically. For example, the central hypothesis of the double filtering by frequency model[37] is that the left hemisphere performs an operation akin to high-pass filtering of the sensory representation of a stimulus whereas the right hemisphere performs low-pass filtering. In the present model, the differential routing of information processed on the short 25–50 ms timescale is equivalent to high-pass filtering, whereas processing on the longer 200–300 ms timescale produces the effect of low-pass filtering. This parallels findings from electrophysiological recordings in marmosets showing two neuronal populations with distinct temporal processing profiles in STG[38].

The present model is also consistent with the proposal suggesting that left auditory cortex specializes in processing stimuli requiring enhanced temporal resolution, whereas right auditory cortex specializes in processing stimuli requiring higher frequency resolution[4]. The output of the analysis on the 25–50 ms timescale will contain more transient features than that resulting from the 200–300 ms analysis, thus manifesting enhanced temporal resolution. The parallel between analysis on the 200–300 ms timescale and enhanced frequency resolution is more subtle and derives from consideration of the minimum duration of a signal necessary to produce a reliable estimate of dynamic pitch and sentential prosody, which is on the order of 200–300 ms. The asymmetric sampling in time (AST) model[3], although based on a different specification of the spatial distribution of short- and long-duration analysis windows, agrees conceptually with the present model and makes similar predictions for right-hemisphere function. For example, AST predicts a rightward lateralization of function for long-duration segments, owing to temporal integration with a time constant of 200–300 ms in the right hemisphere, and a leftward lateralization for short segments owing to a time constant of 25–50 ms in the left hemisphere. Although a rightward lateralization for-long duration stimuli was observed in the present study, the corresponding leftward lateralization was not observed. That shortcoming of AST—the localized processing posited in each hemisphere—is circumvented here by attributing lateralization to differential routing of information.

Summary

We propose a model that accounts for the (i) bilateral temporal sensitivity of auditory cortices (ii) differential sensitivity to local spectrotemporal structure in STG, and (iii) hemispheric lateralization in STS. It is characterized by distributed and hierarchical processing on multiple timescales[3,12,39], because different timescales carry distinct and functionally relevant information about a signal. We hypothesize that there exist (at least) two timescales in STG pertinent to cortical auditory processing, 25–50 ms and 200–300 ms, and that neuronal populations corresponding to each timescale differentially target STS, with the right hemisphere receiving afferents carrying information processed on the long timescale and the left hemisphere those resulting from processing on the short timescale.

The observed symmetries and asymmetries follow from the interaction of the physiological properties of the neuronal ensembles that mediate the analysis of auditory signals with the spectrotemporal properties of these signals. The model that emerges is one of a progressive elaboration of the representation of an acoustic signal by a distributed and hierarchical cortical architecture in which sensory features extracted by earlier stages of processing facilitate the subsequent construction of stable perceptual codes. Such a model provides a unifying and neurophysiologically grounded account of early auditory cortical processing.

METHODS

Subjects. Fifteen (9 female, 6 male) right-handed subjects aged 18–40 participated in the study. All subjects were free of neurological or medical illnesses, had normal structural MRI and audiometric examinations, gave written informed consent and were paid for their participation.

Stimuli. A total of 13 stimulus conditions (6 tonal TN, 6 frequency-modulated FM and 1 control CN) were created by concatenating narrow-band noise segments consisting of a sum of 50 sinusoids with randomized amplitude phase and frequency. Segment bandwidth spanned a half-octave frequency range of 1,000–1,500 Hz and had a bandwidth of 125 Hz, corresponding to a typical second formant[16] and remaining within the critical band at that frequency[17]. Two types of segments were used, one in which the frequency remained constant throughout the segment (**Fig. 1b**; TN) and one in which frequency was swept linearly upward or downward randomly (**Fig. 1a**; FM). For the TN stimuli, the center frequency was drawn from a uniform distribution spanning the half-octave range. For the FM stimuli, all of the individual segments swept up or down over the same frequency range. For each of the two segment types, six mean segment durations of 12, 25, 45, 85, 160 and 300 ms were drawn from a Gaussian distribution with the tails 'clipped' so as not to overlap each other, and with a standard deviation equal to 20% of the mean (**Fig. 1f**). A single control stimulus (CN) was constructed from a single 9-s TN segment with center frequency in the middle of the half-octave range. The amplitudes of all 13 stimuli were adjusted so that all had equal RMS power (**Fig. 1d**). When collapsed over the entire 9-s stimulus duration, all conditions had a nearly equal spectral profile (**Fig. 1e**).

Procedure. Presentation software (version 0.43, Neurobehavioral Systems) running on a PC controlled stimulus delivery by synchronizing each trial to a series of TTL pulses produced during image acquisition. Sound was delivered binaurally to a Commander XG audio system (Resonance Technology). The electrostatic headphones provide ~30 dB of sound attenuation and reduce the ambient MRI scanner noise. The sound level was set to 80 dB SPL.

During behavioral testing, participants were presented a set of stimuli via Presentation identical to those used during imaging except with a segment duration range increased to 10–600 ms (comprising 11 conditions). Subjects were instructed to classify the perceptibility of the individual segments in a single-interval, two-alternative forced-choice task as either perceptible or not perceptible by pressing the appropriate button on a computer keyboard.

Image acquisition. All images were obtained from a 1.5-T GE Signa scanner (GE Medical Systems) equipped with a standard quadrature head-coil. Functional images were collected using a single-shot echoplanar pulse sequence (TE, 40 ms; TR, 11.4 s; flip angle, 90°). Subjects listened passively to a total of 24 trials of each of the 13 stimuli in 8 blocks. Blocks consisted of 39 pseudo-randomized trials during which each condition occurred three times, producing a total of 312 images per subject. Two images were added at the beginning of each block to allow the hemodynamic response to equilibrate, and subsequently discarded from further analysis. The 11.4 s 'clustered' volume acquisition consisted of 9 s of stimulus presentation followed by 2.4 s of slice acquisition to minimize contamination from artifacts induced by scanner[14,15]. Thus, subjects never heard (gradient coil) scanner noise while listening to the stimuli. A total of 24 slices were acquired to provide whole-brain coverage. Subject head movement was minimized by wrapping the head with a vacuum pillow.

Image analysis. Image preprocessing and statistical analysis were performed using SPM99b software. Anatomical segmentation was performed in Matlab version 6 (Mathworks) based on labeled voxels provided by the Talairach

Daemon, version 1.1 (Research Imaging Center, University of Texas Health Science Center). Image preprocessing consisted of realigning the image time series to the first image to correct for subject movement, smoothing via convolution with a Gaussian kernel (8 mm full-width half-maximum), high-pass filtering to remove slow drifts in signal intensity, and re-sampling from a grid size of $3.75 \times 3.75 \times 5$ mm (24 slices) to $4 \times 4 \times 4$ mm (34 slices). In the fixed-effects cohort analysis, categorical contrasts were constructed between each of the six TN and FM conditions and the CN stimulus, resulting in a *t*-statistic for each voxel and thresholded at $P < 0.05$ (corrected via Gaussian random fields for multiple comparisons). The thresholded *t*-score maps (SPMs) were then spatially normalized to the MNI T1 template. The coordinates of all suprathreshold voxels in MNI space were converted to Talairach coordinates and uploaded to the Talairach Daemon for anatomical classification. Although the Talairach Daemon conflates areas STS and MTG at some voxels, visual inspection of the co-registered EPI (echo planar imaging) and anatomical data showed that the activation occurred primarily in STS. Thus, we refer to these voxels simply as STS. In the subsequent single-subject ROI analysis, the effect of segment type and hemisphere were assessed. Mean voxel strength was computed by extracting all voxels from the categorical contrasts within a region defined by the union of suprathreshold voxels from the cohort analysis. Voxel strengths were then submitted to a three-way, full-factorial, repeated-measures ANOVA with hemisphere, segment type (TN or FM) and segment duration as factors.

Surface maps in **Figure 2** were created by converting the cohort activation from MNI space to Talairach space in AFNI and overlaying it on the N27 Brain, also warped to Talairach space, using SUMA.

URLs. AFNI, http://afni.nimh.nih.gov/afni; N27 Brain, http://www.bic.mni.mcgill.ca and UCLA http://www.loni.ucla.edu; SUMA, http://afni.nimh.nih.gov/afni/suma.

Note: Supplementary information is available on the Nature Neuroscience website.

ACKNOWLEDGMENTS

We thank P. Bandettini, J. Fritz, A.-L. Giraud, A. Martin and J. Rauschecker for insightful critical comments; F. Husain for help with experimental setup; and K.M. Boemio for her continued and continuous encouragement. A.B. and D.P. were supported by US National Institutes of Health R01 DC05660 to D.P. During the preparation of the manuscript, D.P. was a fellow at the Wissenschaftskolleg zu Berlin and the American Academy Berlin.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Received 20 December 2004; accepted 25 January 2005.
Published online at http://www.nature.com/natureneuroscience/

1. Hall, D.A., Hart, H.C. & Johnsrude, I.S. Relationships between human auditory cortical structure and function. *Audiol. Neurootol.* **8**, 1–18 (2003).
2. Zatorre, R., Belin, P. & Penhune, V. Structure and function of auditory cortex: music and speech. *Trends Cogn. Sci.* **6**, 37–46 (2002).
3. Poeppel, D. The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time'. *Speech Commun.* **41**, 245–255 (2003).
4. Zatorre, R. & Belin, P. Spectral and temporal processing in human auditory cortex. *Cereb. Cortex* **11**, 946–953 (2001).
5. Binder, J.R. *et al.* Human temporal lobe activation by speech and nonspeech sounds. *Cereb. Cortex* **10**, 512–528 (2000).
6. Scott, S.K. & Johnsrude, I.S. The neuroanatomical and functional organization of speech perception. *Trends Neurosci.* **26**, 100–107 (2003).
7. Hickok, G. & Poeppel, D. Towards a functional neuroanatomy of speech perception. *Trends Cogn. Sci.* **4**, 131–138 (2000).
8. Johnsrude, I.S., Penhune, V.B. & Zatorre, R.J. Functional specificity in the right human auditory cortex for perceiving pitch direction. *Brain* **123**, 155–163 (2000).
9. Scott, S.K., Blank, C.C., Rosen, S. & Wise, R.J. Identification of a pathway for intelligible speech in the left temporal lobe. *Brain* **123**, 2400–2406 (2000).
10. Gandour, J. *et al.* A cross-linguistic FMRI study of spectral and temporal cues underlying phonological processing. *J. Cogn. Neurosci.* **14**, 1076–1087 (2002).
11. Giraud, A-L. *et al.* Representation of the temporal envelope of sounds in the human brain. *J. Neurophysiol.* **84**, 1588–1598 (2000).
12. Shamma, S. On the role of space and time in auditory processing. *Trends Cogn. Sci.* **5**, 340–348 (2001).
13. Hall, D.A. *et al.* Spectral and temporal processing in human auditory cortex. *Cereb. Cortex* **12**, 140–149 (2002).
14. Hall, D.A. *et al.* 'Sparse' temporal sampling in auditory fMRI. *Hum. Brain Mapp.* **7**, 213–223 (1999).
15. Edmister, W.B., Talavage, T.M., Ledden, P.J. & Weisskoff, R.M. Improved auditory cortex imaging using clustered volume acquisitions. *Hum. Brain Mapp.* **7**, 89–97 (1999).
16. Stevens, K.N. *Acoustic Phonetics* (MIT Press, Cambridge, Massachusetts, USA, 1998).
17. Moore, B.C.J. in *Human Psychophysics* (eds. Yost, W.A., Popper, A.N. & Fay, R.R.) (Springer, New York, 1993).
18. Hackett, T.A., Preuss, T.M. & Kaas, J.H. Architectonic identification of the core region in auditory cortex of macaques, chimpanzees, and humans. *J. Comp. Neurol.* **441**, 197–222 (2001).
19. Griffiths, T.D., Buchel, C., Frackowiak, R.S. & Patterson, R.D. Analysis of temporal structure in sound by the human brain. *Nat. Neurosci.* **1**, 422–427 (1998).
20. Harms, M.P. & Melcher, J.R. Sound repetition rate in the human auditory pathway: representations in the waveshape and amplitude of fMRI activation. *J. Neurophysiol.* **88**, 1433–1450 (2002).
21. Wang, X., Lu, T. & Liang, L. Cortical processing of temporal modulations. *Speech Commun.* **41**, 107–121 (2003).
22. Wessinger, C.M. *et al.* Hierarchical organization of the human auditory cortex revealed by functional magnetic resonance imaging. *J. Cogn. Neurosci.* **13**, 1–7 (2001).
23. Yost, W.A. Auditory image perception and analysis: the basis for hearing. *Hear. Res.* **56**, 8–18 (1991).
24. Yabe, H. *et al.* Organizing sound sequences in the human brain: the interplay of auditory streaming and temporal integration. *Brain Res.* **897**, 222–227 (2001).
25. Winkler, I., Reinikainen, K. & Naatanen, R. Event-related brain potentials reflect traces of echoic memory in humans. *Percept. Psychophys.* **53**, 443–449 (1993).
26. Sussman, E., Winkler, I., Ritter, W., Alho, K. & Naatanen, R. Temporal integration of auditory stimulus deviance as reflected by the mismatch negativity. *Neurosci. Lett.* **264**, 161–164 (1999).
27. Zwislocki, J. Theory of temporal auditory summation. *J. Acoust. Soc. Am.* **32**, 1046–1060 (1960).
28. Kaas, J.H. & Hackett, T.A. Subdivisions of auditory cortex and processing streams in primates. *Proc. Natl Acad. Sci. USA* **97**, 11793–11799 (2000).
29. Geschwind, N. & Levitsky, W. Human brain: left-right asymmetries in temporal speech region. *Science* **161**, 186–187 (1968).
30. Galuske, R.A., Schlote, W., Bratzke, H. & Singer, W. Interhemispheric asymmetries of the modular structure in human temporal cortex. *Science* **289**, 1946–1949 (2000).
31. Jäncke, L., Wustenberg, T., Scheich, H. & Heinze, H.J. Phonetic perception and the temporal cortex. *Neuroimage* **15**, 733–746 (2002).
32. Näätänen, R. *et al.* Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature* **385**, 432–434 (1997).
33. Palva, S. *et al.* Distinct gamma-band evoked responses to speech and non-speech sounds in humans. *J. Neurosci.* **22**, RC211 (2002).
34. Schwartz, J. & Tallal, P. Rate of acoustic change may underlie hemispheric specialization for speech perception. *Science* **207**, 1380–1381 (1980).
35. Divenyi, P.L. & Robinson, A.J. Nonlinguistic auditory capabilities in aphasia. *Brain Lang.* **37**, 290–326 (1989).
36. Meyer, M., Alter, K., Friederici, A.D., Lohmann, G. & von Cramon, D.Y. FMRI reveals brain regions mediating slow prosodic modulations in spoken sentences. *Hum. Brain Mapp.* **17**, 73–88 (2002).
37. Ivry, R.B. & Robertson, L.C. *The Two Sides of Perception* (Bradford Books, MIT Press, Cambridge, Massachusetts, 1997).
38. Lu, T., Liang, L. & Wang, X. Temporal and rate representations of time-varying signals in the auditory cortex of awake primates. *Nat. Neurosci.* **4**, 1131–1138 (2001).
39. Smith, Z.M., Delgutte, B. & Oxenham, A.J. Chimaeric sounds reveal dichotomies in auditory perception. *Nature* **416**, 87–90 (2002).