

Topographic Mapping of a Hierarchy of Temporal Receptive Windows Using a Narrated Story

Yulia Lerner, Christopher J. Honey, Lauren J. Silbert, and Uri Hasson

Department of Psychology and the Neuroscience Institute, Princeton University, Princeton, New Jersey 08540-1010

Real-life activities, such as watching a movie or engaging in conversation, unfold over many minutes. In the course of such activities, the brain has to integrate information over multiple time scales. We recently proposed that the brain uses similar strategies for integrating information across space and over time. Drawing a parallel with spatial receptive fields, we defined the temporal receptive window (TRW) of a cortical microcircuit as the length of time before a response during which sensory information may affect that response. Our previous findings in the visual system are consistent with the hypothesis that TRWs become larger when moving from low-level sensory to high-level perceptual and cognitive areas. In this study, we mapped TRWs in auditory and language areas by measuring fMRI activity in subjects listening to a real-life story scrambled at the time scales of words, sentences, and paragraphs. Our results revealed a hierarchical topography of TRWs. In early auditory cortices (A1+), brain responses were driven mainly by the momentary incoming input and were similarly reliable across all scrambling conditions. In areas with an intermediate TRW, coherent information at the sentence time scale or longer was necessary to evoke reliable responses. At the apex of the TRW hierarchy, we found parietal and frontal areas that responded reliably only when intact paragraphs were heard in a meaningful sequence. These results suggest that the time scale of processing is a functional property that may provide a general organizing principle for the human cerebral cortex.

Introduction

Space and time are two fundamental properties of our physical and psychological realms. Much is known about the integration of information across space within the visual system (Sereno et al., 1995; DeYoe et al., 1996; Malach et al., 2002; Larsson et al., 2006; Wandell et al., 2007; Silver and Kastner, 2009), but little is known about the integration of information over seconds and minutes of time. It is a basic organizing principle of the visual system that neurons along cortical visual pathways have increasingly large spatial receptive fields (Hubel, 1988); neurons in higher-level visual areas receive input from many neurons with smaller receptive fields in early visual areas, and thereby aggregate information across space. Real-world events, however, occur not only over extended regions of space but also over extended periods of time, suggesting that an analogous gradient of scaling and selectivity may exist in the temporal domain.

We recently demonstrated that the reliability of brain responses varies differentially across visual areas as a function of the temporal structure of a silent movie sequence (Hasson et al., 2008). Drawing an analogy with the spatial receptive field (SRF), we define the temporal receptive window (TRW) of a cortical microcircuit as the length of time before a response during which sensory information may affect that response. Our findings in the

visual system were consistent with the hypothesis that TRWs increase as one moves from low-level (sensory) to high-level (perceptual and cognitive) areas.

To further test this hypothesis and to assess its domain generality, we mapped the topographic organization of TRWs within the auditory system. We predicted that, as in the visual system, the TRWs increase in a gradual topographic manner from early auditory areas to high-order language areas and, moreover, that frontal areas will exhibit the capacity to accumulate information over the longest window of time.

The auditory system can spectrally decompose (Scheich, 1991) a diversity of natural sensations, but is also specialized (Telkemeyer et al., 2009) toward the detection of linguistically salient sound patterns and language processing in general. Furthermore, there is a natural correspondence between distinct time scales of sensory input and separable linguistic units such as words, sentences, and paragraphs. The latter (paragraph and longer) time scale determines the context (Xu et al., 2005) or narrative framework (Fletcher et al., 1995; Ferstl and von Cramon, 2002; Ferstl et al., 2005; Whitney et al., 2009) within which much natural communication takes place.

To assess the time scale of processing within and beyond auditory cortex, we measured the reliability of neural responses evoked by parametrically scrambled versions of a real-life, 7 min story. We scrambled the story at the word level, sentence level, and paragraph level, and also played it backward. All scrambled versions were reorderings of the same sound segments, varying only in the coherence of their temporal structure. Although the stimuli were segmented at natural linguistic boundaries for the purposes of scrambling, this study was not designed to localize linguistic or semantic units such as a words area or a sentences

Received July 14, 2010; revised Nov. 15, 2010; accepted Dec. 22, 2010.

This work was supported by National Institutes of Health Grant R21-DA024423 (to C.H.). We thank Simon Garrod, Adele Goldberg, Anne Gilman, David Heeger, Ifat Levy, Rafael Malach, and Nava Rubin for very fruitful discussions. Special thank to Mikhail Katkov for programming assistance.

Correspondence should be addressed to Prof. Uri Hasson, 3-C-13 Green Hall, Department of Psychology, Princeton University, Princeton, NJ 08540-1010. E-mail: hasson@princeton.edu.

DOI:10.1523/JNEUROSCI.3684-10.2011

Copyright © 2011 the authors 0270-6474/11/312906-10\$15.00/0

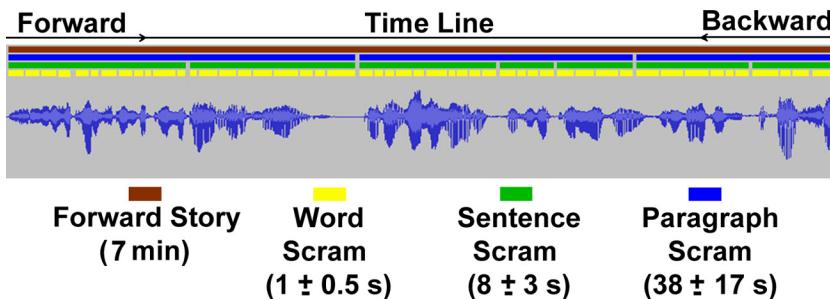


Figure 1. Examples of stimuli. A schematic of the stimuli used in the experiment. Forward (brown), backward, and scrambled (scram) versions of a real-life, 7 min story were used. To generate scrambled versions of the story, natural segments, defined by the end points of each word, sentence, and paragraph, were randomly shuffled at each of the following time scales: words (yellow), sentences (green), and paragraphs (blue).

area. Rather, by comparing neural responses to these differently scrambled stimuli, we aimed to characterize one property of the underlying computational process in each brain area: its sensitivity to information that arrived at differing points in the past.

To assess the reliability of the responses to each stimulus in a given brain region, we measured, across individuals, the correlation of the blood oxygenation level-dependent (BOLD) signals evoked in that area. The results indicated that the reliability of responses varies systematically across brain areas as a function of temporal structure. The response reliability in early auditory areas was not affected by the temporal ordering of events. In contrast, the reliability of responses in higher brain areas increased gradually in correspondence with the length of coherent temporal structures in the stimulus. This hierarchical temporal topography corroborates and extends previous findings in the visual modality. Further, it supports the notion that the sensitivity to information across time, or TRW, is a general functional property and one that may help to bridge our understanding of perceptual, cognitive, and working memory processes in the human brain.

Materials and Methods

Subjects

Fifteen subjects (ages 20–36 years), participated in the fMRI study. Conditions in which the head motion were >1 mm or where the signal was corrupted were discarded from the analysis, and additional subjects were scanned until data from 11 subjects were collected for each condition. Overall, seven subjects participated in all five conditions, one subject in four conditions, four subjects in three conditions, and three subjects in one or two conditions. Procedures were approved by the Princeton University Committee on Activities Involving Human Subjects. All subjects had normal hearing and provided written informed consent.

MRI acquisition

Subjects were scanned in a 3T head-only MRI scanner (Allegra; Siemens). A custom radio-frequency coil was used for the structural scans (NM-011 transmit head coil; Nova Medical). For fMRI scans, 300 volumes were acquired using a T2*-weighted echo planar imaging (EPI) pulse sequence [repetition time (TR), 1500 ms; echo time (TE), 30 ms; flip angle, 75°], each volume comprising 25 slices of 3 mm thickness with 1 mm gap (in-plane resolution, 3×3 mm 2). Slice acquisition order was interleaved. In addition, a set of 160 T1-weighted high-resolution ($1 \times 1 \times 1$ mm 3) anatomical images of the same orientation as the EPI slices were acquired for each subject with a magnetization-prepared rapid-acquisition gradient echo (MP-RAGE) pulse sequence [TR, 2500 ms; TE, 4 ms; slice thickness, 1 mm; no gap; in-plane resolution, 1×1 mm 2 ; field of view, 256 mm 2] and used for cortical segmentation and three-dimensional (3D) reconstruction. To minimize head movement, subjects' heads were stabilized with foam padding. Stimuli were presented using Psychophysics toolbox (Brainard, 1997; Pelli, 1997). MRI-compatible head-

phones (MR Confon) were fitted to provide considerable attenuation to the scanner noise and to present the audio stimuli to the subjects.

Stimuli and experimental design

Main experiment. Stimuli for the experiment were generated from a 7 min real-life story ("Pie-man," told by Jim O'Grady) recorded at a live storytelling performance ("The Moth" storytelling event, New York City). Subjects listened to the whole story from beginning to end (intact forward story), as well as to the story presented waveform-reversed in time (backward story). Subjects also listened to scrambled stimuli, which were created by randomly shuffling segments of the intact story. The story was segmented manually by identifying the end points of each word, sentence, and paragraph.

A segment of two short words was assigned in cases where we could not separate those adjacent words. Following segmentation, the intact story was scrambled at three time scales: short (608 words, 0.7 ± 0.5 s each), intermediate (69 sentences, 7.7 ± 3.5 s each), and long (11 paragraphs, 38.1 ± 17.6 s each). Laughter and applause were classified as single word events (4.4% of the words). Twelve seconds of neutral music and 3 s of silence preceded and 15 s of silence followed each playback in all conditions. These music and silence periods were discarded from all analyses. A typical session was comprised of five runs, each consisting of the presentation of one condition. Presentation order was pseudorandomized across subjects. Attentive listening to the story was confirmed using a simple questionnaire at the conclusion of the experiment.

Localizer experiment. To localize areas that responded reliably to natural stories, an independent, 15 min, real-life story was played to each subject. A transcript of the story is provided in Stephens et al. (2010). An intersubject correlation analysis (see below) was used to identify areas that responded reliably to this localizer story. This reliability map was used to define independent regions of interest (ROIs) that exhibited reliable responses to natural, real-life stories (see ROI analysis, below).

Data analysis

Preprocessing. fMRI data were reconstructed and analyzed with the BrainVoyager QX software package (Brain Innovation) and with in-house software written in MATLAB (MathWorks). Preprocessing of functional scans included intrasession 3D motion correction, slice scan time correction, linear trend removal, and high-pass filtering (three cycles per experiment). Spatial smoothing was applied using a Gaussian filter of 6 mm full-width at half-maximum value. The cortical surface was reconstructed from the 3D MP-RAGE anatomical images using standard procedures implemented in the BrainVoyager software. The complete functional dataset was transformed to a 3D Talairach space (Talairach and Tournoux, 1988) and projected on a reconstruction of the cortical surface.

Intersubject correlation analysis. The central results of this study are derived from the intersubject correlation analysis. This analysis provides a measure of the reliability of the responses to a temporally complex stimulus, such as a story, by comparing the BOLD response time courses across different subjects [intersubject correlation (inter-SC)]. In another variant of this method, one can calculate the reliability of responses within an individual by comparing the BOLD responses to the same stimulus repeatedly presented within the same individual [intrasubject correlation (intra-SC)] (Golland et al., 2007; Hasson et al., 2009). Intra-SC and inter-SC are conceptually distinct metrics; an individual could exhibit an idiosyncratic response that is the same on every presentation, which would be show up as reliable when assessed with intra-SC but unreliable when assessed with inter-SC (Hasson et al., 2009). However, in a previous study, we found very similar time scale topographies within (intra-) and between (inter-) subjects in the visual system (Hasson et al., 2008). In this study, we focused solely on the reliability of responses across subjects (inter-SC). The inter-SC method differs from conventional fMRI data analysis methods in that it circumvents the need to specify a model for the neuronal processes in any given brain region during

movie watching. Instead, the inter-SC method uses the subject's brain responses to naturalistic stimuli (for example, a narrated story) as a model to predict brain responses within other subjects.

Correlation coefficients were calculated on a voxel-by-voxel basis (in Talairach space) within each condition (forward, backward, words, sentences, and paragraphs) by comparing the responses across all listeners (inter-SC). The analysis produced, for each stimulus condition, an average inter-SC map, which was constructed as explained below.

First, at every voxel, the Pearson product-moment correlation coefficient ρ_k of subject k was computed as follows:

$$\rho_k = \rho(r_k, \bar{r}) = \frac{r_k(t) \cdot \bar{r}(t)}{\sqrt{(r_k(t) \cdot r_k(t))(\bar{r}(t) \cdot \bar{r}(t))}},$$

where $r_k(t)$ is the mean-subtracted response time course of a voxel to the stimulus presentation for a subject k , $\bar{r}(t) = \sum_{i \neq k} r_i(t)$ is the response time course averaged across all subjects except k , and $r_k(t) \cdot \bar{r}(t)$ is the inner product of $r_k(t)$ and $\bar{r}(t)$. Next, to create a multisubject map, the average correlation $\rho = \sum \rho(r_k, \bar{r})$ was computed at every voxel. Statistical significance of each ρ value was computed using a bootstrapping procedure. For every empirical time course in every voxel, 1000 bootstrap time series were generated using a phase-randomization procedure. Phase randomization was performed by fast Fourier transforming the signal, $\bar{r}(t)$, randomizing the phase of each Fourier component, and then inverting the Fourier transformation. This procedure leaves the power spectrum of the signal unchanged but removes temporal alignment of the signals. Using these bootstrap time courses, a null distribution of the correlation values was determined for each subject, k , in each voxel, and these were combined to produce a null distribution of the average correlation, $\bar{\rho}$. The null distributions of average correlations generated in this way at each voxel were found to follow an approximately normal distribution and, therefore, they were fitted at each voxel by a normal distribution. The p values of the empirical average correlations were computed by comparison with these null distributions.

Finally, to correct for multiple comparisons, we applied the Benjamini–Hochberg–Yekutieli false-discovery procedure, which controls the false discovery rate (FDR) under assumptions of dependence (Benjamini and Hochberg, 1995; Benjamini and Yekutieli, 2001). More specifically, after sorting voxels in order of ascending p value, the significantly correlated across-subject voxels were defined to be the first k , where k is the largest integer, such that:

$$P_k \leq \frac{k}{M \cdot c(m)} q,$$

where P_k is the p value of t for voxel k , M is the number of voxels, $c(m) = 1$, assuming posi-

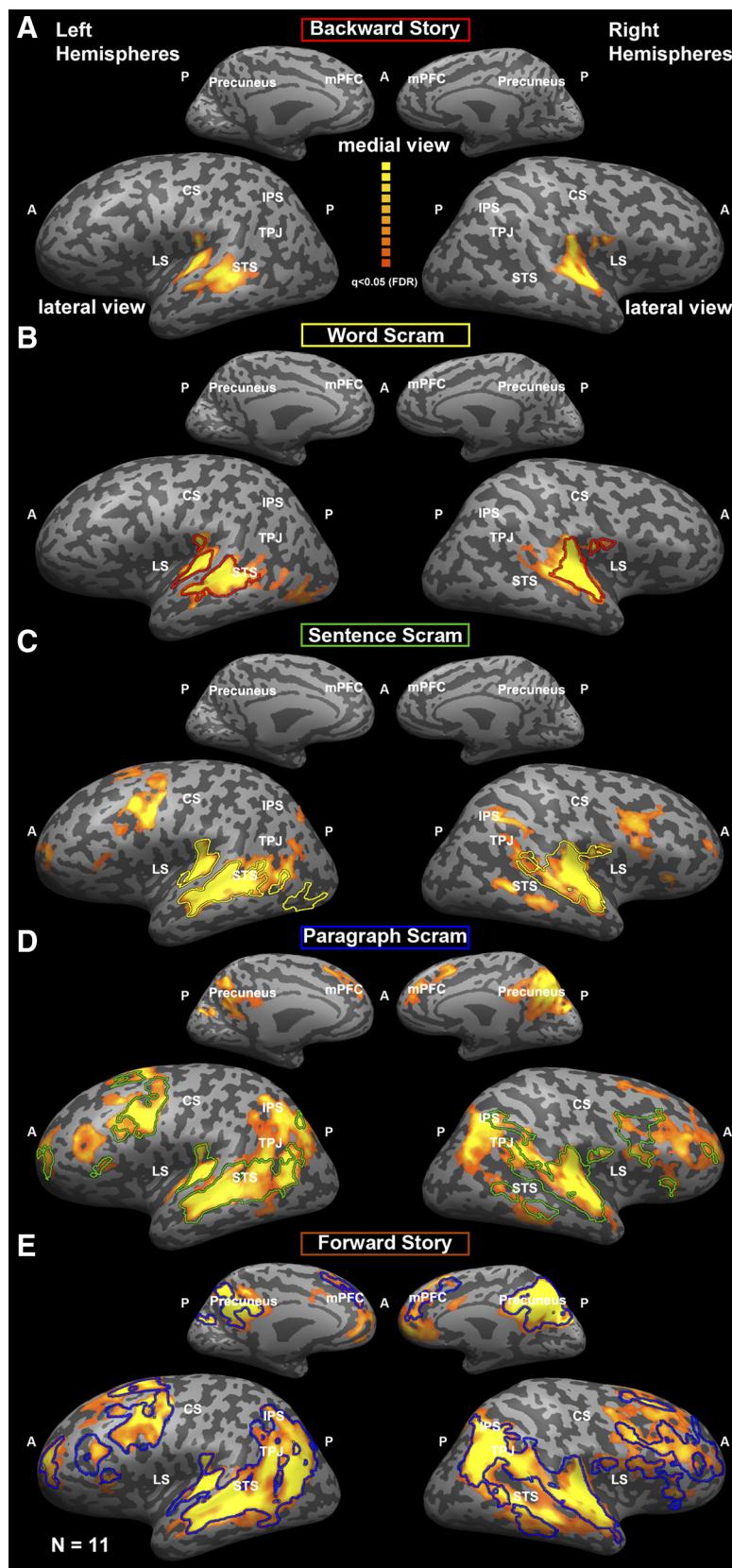


Figure 2. Effect of scrambling at different time scales. Maps of the reliability of responses across subjects were computed separately for each condition and superimposed on an inflated brain shown in lateral and medial views. The maps illustrate the extent of the reliable responses at sequentially higher levels of temporal coherence (**A**), word scrambled (**B**), sentence scrambled (**C**), paragraph scrambled (**D**), and intact forward story (**E**) and each map subsumes the borders of the significantly reliable responses at the previous level. LS, Lateral sulcus; STS, superior temporal sulcus; CS, central sulcus; IPS, intraparietal sulcus; A, anterior; P, posterior.

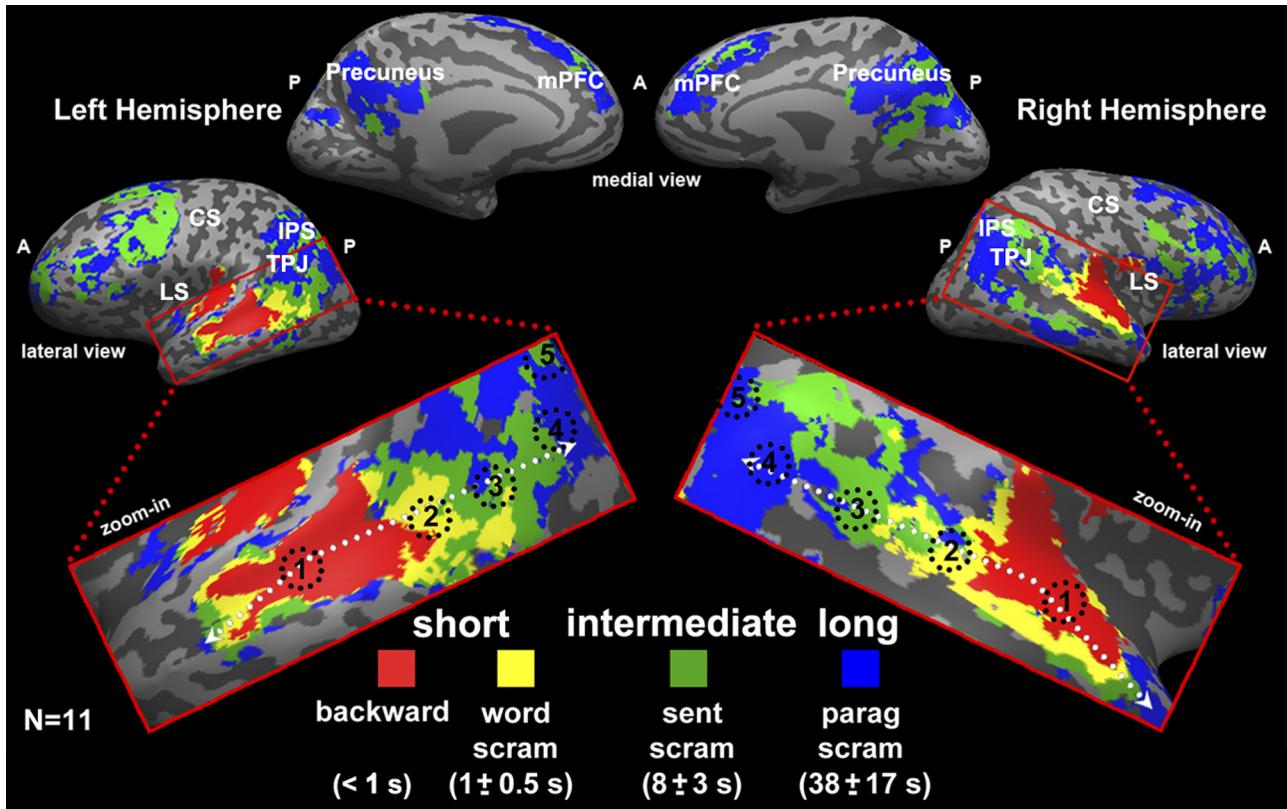


Figure 3. Hierarchical topography of temporal receptive windows. Response reliability (inter-SC) to auditory narratives as a function of temporal structures average across all subjects. Circled numbers correspond to graphs in Fig. 5. A voxel was assigned the label “backward” (red) when it was significant in all inter-SC maps (Fig. 5, ROI 1). A voxel was assigned the label “word scramble” (yellow) when it was significant in all inter-SC maps except backward (Fig. 5, ROI 2). Similarly, a voxel was assigned the label “sentence scramble” (green) when it was significant in all inter-SC maps except word scramble and backward (Fig. 5, ROI 3). A voxel was assigned the label “paragraph scramble” (blue) when it was significant only in the paragraph scrambled inter-SC maps (Fig. 5, ROI 4). Circled numbers indicate the approximate location of the ROIs along the A1-TPJ axis used in the ROI analysis. LS, Lateral sulcus; STS, superior temporal sulcus; CS, central sulcus; IPS, intraparietal sulcus; A, anterior; P, posterior; scram, scrambled.

tively correlated t values across voxels, and using a false discovery threshold $q = 0.05$.

ROI analysis. Widespread areas of cortex responded reliably ($q < 0.05$, FDR corrected) when subjects listened to an independent story (see Localizer experiment, above). The inter-SC map obtained for the main story overlapped with the inter-SC map obtained for the localizer story (compare Fig. 2E with supplemental Fig. 1A, available at www.jneurosci.org as supplemental material). To sample the spatial axis from area A1 toward the temporal-parietal junction (TPJ), we defined a set of ROIs by manually partitioning the extent of the significantly responsive voxels in the independent localizer map into five approximately equally sized adjacent subregions (labeled 1 through 5; Talairach coordinates: ROI 1: $\pm 53, -19, 5$; ROI 2: $\pm 58, -41, 0$; ROI 3: $\pm 58, -53, 0$; ROI 4: $\pm 56, -50, 16$; ROI 5: $\pm 54, -76, 17$). Two additional ROIs in the precuneus and medial prefrontal cortex (mPFC) were defined from the data in localizer experiment based on their corresponding anatomical landmarks.

Additionally, we continuously sampled all voxels along the temporal-parietal axis in the extension of reliable responses in localizer experiment (supplemental Fig. 1A, available at www.jneurosci.org as supplemental material). Then, we calculated a TRW index with the following formula: TRW index = $(\rho_{\text{Full Story}} + \rho_{\text{Paragraphs}} + \rho_{\text{Sentences}})/3 - (\rho_{\text{Words}} + \rho_{\text{Reverse}})/2$. Values closer to 0 indicate that the voxel responded equally to all conditions regardless of the level of scrambling, positive values indicate that the responses were higher for conditions with longer temporal coherence (e.g., paragraphs and sentences) than conditions with shorter temporal coherence (e.g., words and backward). The analysis (supplemental Fig. 1B, available at www.jneurosci.org as supplemental material) is equivalent to performing the ROI analysis independently for each voxel along

the temporal-parietal axis, and thus complements the main analysis in which larger and less proximate ROIs are used.

Response amplitude. Due to the nature of our experimental design (continuous real-life stimuli), we cannot directly assess changes in response amplitude relative to a blank baseline. To estimate response amplitude, we quantified the SD of the percentage BOLD signal change time courses. This measure provides a proxy for the overall signal modulations because small signal fluctuations should cause lower SD and larger fluctuations should cause higher SD. The SD was assessed independently for each ROI and each condition, first within each subject and then averaged across subjects.

Unscrambling procedure. The procedure was performed by first segmenting the BOLD responses for each sentence in the sentences-scrambling condition, and for each paragraph in the paragraphs-scrambling condition. The segmentation took the hemodynamic response function delay into account and was performed on the average signal across all subjects within a condition. Due to the hemodynamic response blurring, this procedure was performed only for sentences longer than 6 s; all segments shorter than this period have been excluded from the dataset of all conditions for all unscrambling analyses. Then, we reordered the neural responses of each sentence or paragraph to match the temporal order of the intact forward condition. Finally, the unscrambled sentences response time courses were correlated with the responses to the intact forward story ($C_{FS:UfS}$) and the unscrambled paragraphs response time courses were correlated with the responses to the intact forward story ($C_{FS:Ufp}$). Finally, we compared the forward and time-reversed backward responses ($C_{rb,f}$). To that end, we flipped the average response to the backward condition in each voxel and then shifted the

time courses by $\Delta t = 5$ s to correct for the hemodynamic delay (Hasson et al., 2008).

Results

To examine changes in brain responses as a function of the time scale at which auditory information was preserved, we parametrically varied the time structure of the audio recording. To scramble the story, we first divided it into natural segments defined by the end points of each word, sentence, and paragraph. Next, the temporal order of the story was randomly shuffled at each of the following time scales: words (0.7 ± 0.5 s), sentences (7.7 ± 3.5 s), and paragraphs (38.1 ± 17.6 s). Finally, to thoroughly disrupt the temporal history of each moment in the stimulus, we reversed the story's audio waveform (backward condition). The experimental design is schematized in Figure 1 (see Materials and Methods for more details).

Response reliability across brain regions for different time scales

The data were analyzed using an inter-SC analysis. First, we mapped the voxel-by-voxel inter-SC within each condition (story, paragraph scrambled, sentence scrambled, word scrambled, and backward) by correlating the fMRI BOLD responses in that voxel across subjects (Fig. 2) (Hasson et al., 2004, 2010). In brain areas where responses are driven primarily by instantaneous features of sensory input, the responses should be reliable in all conditions, regardless of temporal scrambling. In contrast, in brain regions where responses depend on sensory information accumulated over several seconds or more, the reliability of the responses should depend on the time scale of scrambling.

Early auditory areas (A1+) responded reliably to all conditions regardless of their temporal structure and exhibited short TRWs (Fig. 3, red). Moving rostrally and caudally from A1+ to higher-level areas, the reliability of responses became increasingly dependent on the level of stimulus-preserved temporal structure (Figs. 2–4). Areas adjacent to A1+ along the superior temporal gyrus exhibited an intermediate TRW (Fig. 3, yellow and green). The longest TRWs were found in the posterior superior temporal sulcus (pSTS), the TPJ, the precuneus, and the frontal cortex (Fig. 3, blue). In these regions, reliable responses were evoked only by the paragraph scrambled and by the intact story conditions (Fig. 2). The topographic organization of TRWs was strikingly consistent across all levels of analysis [i.e., individual maps (Fig. 4), group map with seven contingent subjects (supplemental Fig. 3, available at www.jneurosci.org as supplemental material), and 11 mixed subjects (Fig. 3)].

A hierarchy of time-scale sensitivity was also observed in an ROI analysis (Fig. 5). To better characterize the time-scale gradient along the superior temporal gyrus, we sampled the time courses of five equally sized ROIs that tiled the A1–TPJ axis (Fig. 3, approximate centroids). ROIs were defined functionally using independent auditory stimuli (for details and Talairach coordinates, see Materials and Methods). Consistent with the voxelwise analysis, we found that early auditory areas (A1+, labeled as ROI 1 on the axis) exhibited high correlations for all conditions, including the backward, words, and paragraphs-scrambling conditions. Such findings suggest that the reliability of responses in this area is not dependent on coherent temporally extended structure. Moving rostrally from A1+ toward the TPJ, the reliability of responses becomes more dependent on intact temporal structure, with ROI 2 responding reliably only to stimuli that preserve the information structure of words and longer, ROI 3 only for sentences and longer, and ROI 4 only for paragraphs and longer. In ROI 5 (the TPJ), the responses were reliable in both the

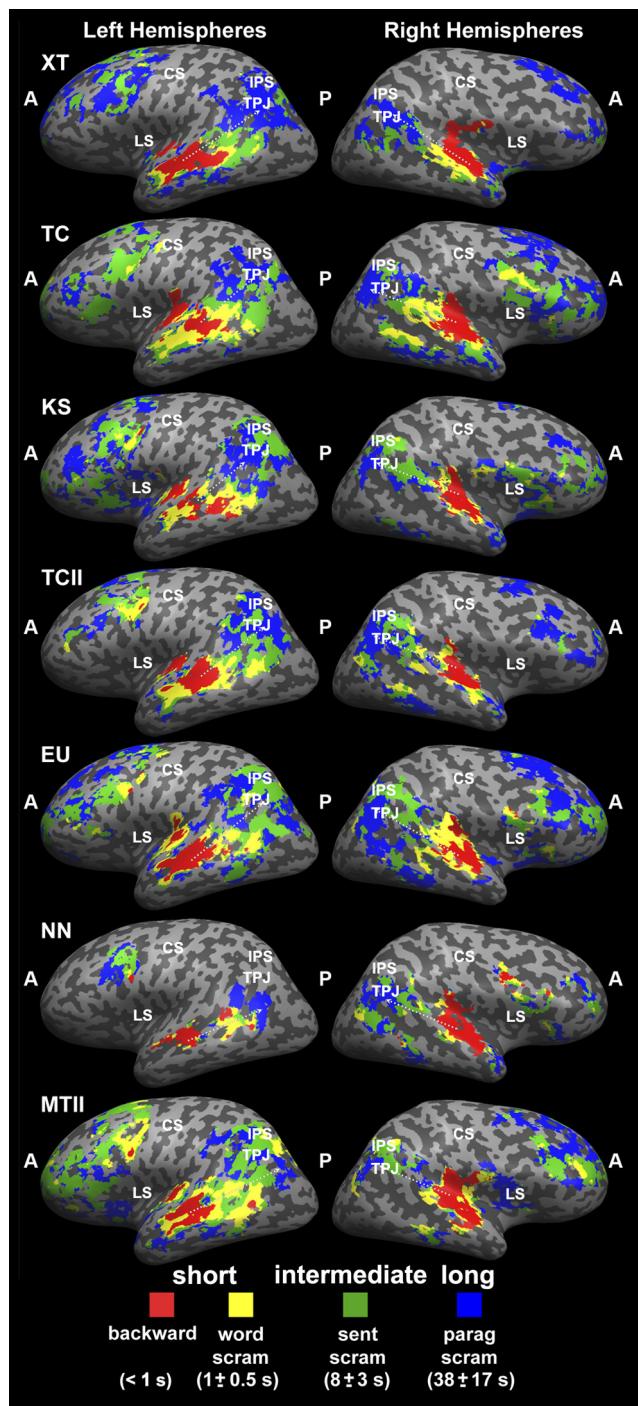


Figure 4. Hierarchical topography of temporal receptive windows in single subjects. Response reliability (inter-SC) to auditory narratives as a function of temporal structures for seven individual subjects, identified by initials. Reliability is assessed by comparing responses of each individual against the average of the responses of the remaining individuals. Figure layout is identical to that in Figure 3. LS, Lateral sulcus; STS, superior temporal sulcus; CS, central sulcus; IPS, intraparietal sulcus; A, anterior; P, posterior; scram, scrambled.

paragraph-scrambled and forward-story conditions, but with significantly greater reliability when the story was presented. To further assess the gradual change in the TRWs, we sampled all voxels along the temporal-parietal axis that exhibited reliable responses in the independent localizer experiment (see Materials and Methods and supplemental Fig. 1, available at www.jneurosci.org as supplemental material). Similar to the voxel-by-

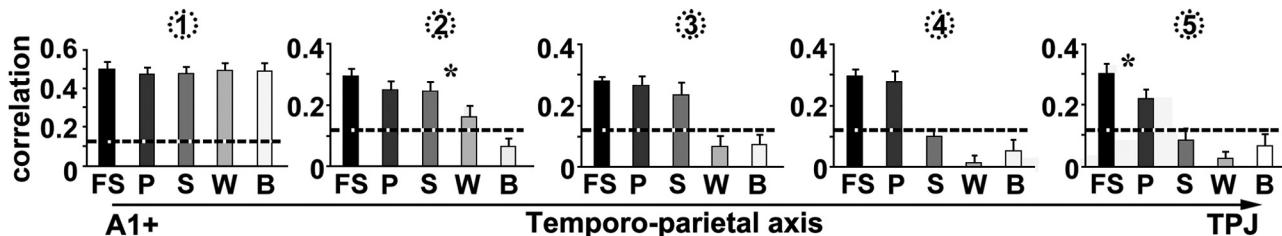


Figure 5. Activation profiles from the ROIs along the A1+–TPJ axis. Inter-SC shown as a function of temporal disruption of the auditory story in several independently defined ROIs. ROIs correspond to circled numbers in Figure 3. Early auditory areas (A1+, ROI 1) exhibited short TRWs, responding reliably to all conditions regardless of the level of temporal scrambling. Areas adjacent to A1+ along the superior temporal gyrus exhibited an intermediate TRW. Here, coherent information at the words (ROI 2), sentences (ROI 3), or longer time scales was necessary to evoke reliable activity. The longest TRWs within this axis were found in the posterior superior temporal sulcus and the TPJ. In these regions, reliable responses were evoked only by the paragraphs condition and the intact story condition (ROIs 4 and 5). FS, Forward story; P, paragraph; S, sentence; W, word; B, backward.

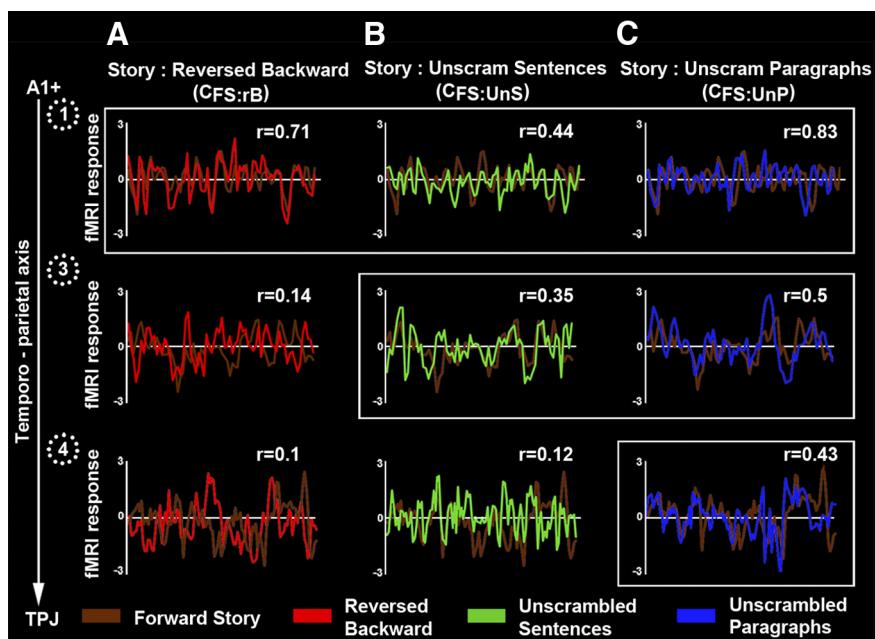


Figure 6. Comparison of responses to the unscrambled (reordered) story with responses to the intact story. Moment-to-moment responses in A1+ evoked by the intact story are similar to the unscrambled backward, unscrambled sentence, and unscrambled paragraph response time courses. In area 3, the time courses to the intact story temporally resemble the unscrambled sentence and unscrambled paragraph time courses. In area 4, only the unscrambled paragraph time course is correlated with the intact story. These findings indicate that the responses within each processing stage are predominantly determined by the temporal structure within its preferred temporal window and are largely insensitive to longer temporal structure.

voxel mapping (Fig. 3) and the ROIs analysis (Fig. 5), we observed a gradual change from areas with short TRWs to areas with long TRWs along the temporal-parietal axis.

We note that scrambling at the word level can slightly decrease the rate of recognition of individual words (Holtzman et al., 1986). Nevertheless, a slight reduction in intelligibility in the words scrambling condition does not compromise the finding of a graded ordered topography of reliability when combining results across all five conditions.

Unscrambling the responses to scrambled stimuli

The TRW hypothesis predicts that a region with a given TRW length will respond similarly to a coherent event of that length regardless of the larger temporal context. For example, if the processing time scale of a region coincides with the sentence time scale, then presenting the same sentences in different orders should not make a difference (i.e., the response pattern evoked by sentence B in such a region will not depend on whether it precedes or

follows sentence A). To test this prediction, we compared the time course evoked by the intact story with unscrambled (reordered) versions of the time courses evoked by the sentence-scrambled and paragraph-scrambled conditions. The unscrambling procedure is feasible only in cases where the length of each segment is longer than the hemodynamically induced temporal blur. We therefore excluded sentences shorter than 6 s (46% of all sentences) from all conditions in this analysis, and did not perform the analysis for the words-scrambled condition (see Materials and Methods for details). In addition, we compared the responses to the intact story with time-reversed responses to the backward story (Hasson et al., 2008).

After unscrambling the brain responses, it was clear that the responses in A1+ were similar across all conditions, regardless of the temporal order. That is, we observed strong correlations when comparing the intact forward story with the time-reversed backward story ($C_{FS:rB}$), with the unscrambled sentences ($C_{FS:UnS}$), and with the unscrambled paragraphs ($C_{FS:UnP}$). The similarities were observed in the raw unscrambled time courses, as well as in the mean inter-SC coefficients (Fig. 6). Thus, consistent with the hypothesis that A1+ has a short TRW, we found that the responses in A1+ are induced by the moment-to-moment auditory input independent of the temporal history.

The TRW within the posterior superior temporal gyrus (ROI 3) (Figs. 3, 5) coincides with the time scale of single sentences. We observed that the unscrambled sentences time course correlated with the responses to the intact story ($C_{FS:UnS}$) (Fig. 6). This indicates that the responses in this ROI are largely unchanged when the sentences are embedded within a meaningful story or presented in a random order. Finally, the TRW within the TPJ (ROI 4) (Figs. 3, 5) coincides with the time scale of the entire paragraph (± 38 s), as the responses to the intact story were only correlated with the unscrambled paragraphs time course ($C_{FS:UnP}$) (Fig. 6).

The longest TRWs were observed in the precuneus and frontal areas

The longest TRWs were observed in the precuneus and mPFC (Fig. 7). These ROIs were defined anatomically using an indepen-

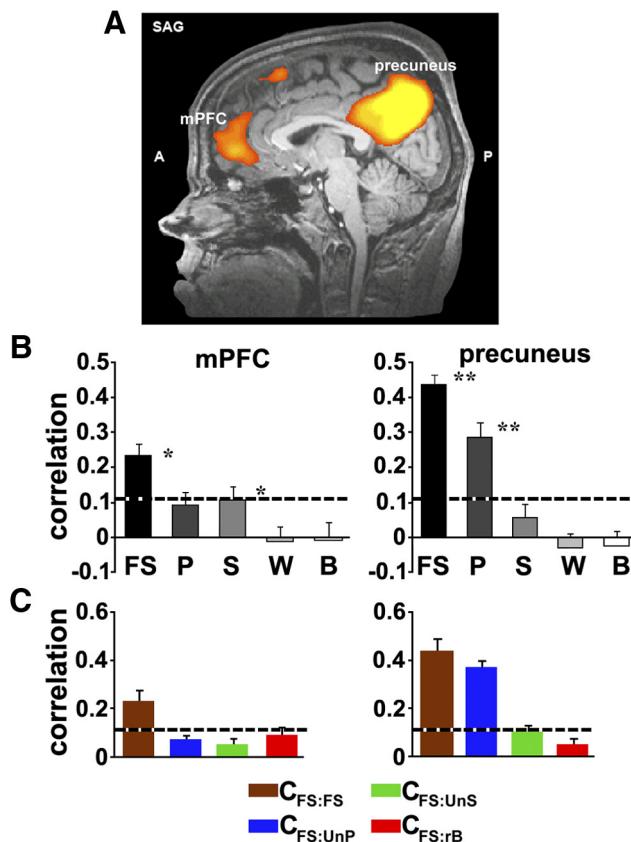


Figure 7. TRW in the high-order mPFC and precuneus. **A**, Inter-SC correlation in the precuneus and mPFC evoked by forward story. **B**, Time course correlation analysis is identical to that shown in Figure 5. **C**, The results of the unscrambling procedure, as applied in areas 1–4 (Fig. 6). In the precuneus, only the unscrambled paragraph time course is correlate with the intact story, confirming that the time scale of processing in this area is relatively long (± 30 s). In the mPFC, the reliable responses during the intact story did not correlate with unscrambled patterns from any of the other conditions. P, Paragraph; S, sentence; W, word; B, backward; FS, Forward story; UnP, unscrambled paragraphs; UnS, unscrambled sentences; rB, reversed backward; A, anterior; P, posterior; SAG, sagittal.

dent localizer. In both areas, we observed reliable responses (high inter-SC) only when presenting subjects with stimuli containing information coherent over long time scales. Moreover, although the precuneus responded reliably to both the paragraph-scrambled and the forward-story conditions, only the forward-story condition evoked reliable responses in the mPFC. Finally, the unscrambling analysis indicates that the responses in the precuneus are similar for both the intact story and for the unscrambled response to the paragraph-scrambled condition ($C_{FS:UnP}$), confirming that the time scale of processing in this area coincides with relatively long time scales (± 30 s). In contrast, the activation time course in the mPFC during the intact story did not correlate with unscrambled patterns from any of the other conditions.

Dissociation between the time scales of processing and response amplitudes

Areas that responded with low reliability to the temporally scrambled and time-reversed audio recordings nevertheless showed high response amplitudes to those same stimuli (supplemental Fig. 2, available at www.jneurosci.org as supplemental material). Response amplitudes were estimated by computing the SD of the responses over time within each ROI. Our analysis revealed that the SDs were indistinguishable for all conditions in all of the brain areas examined (supplemental Fig. 2). The mea-

surement of response reliability can therefore uncover effects that response amplitudes alone do not reveal. Disrupting temporal order had no effect on BOLD response amplitudes, even in brain areas in which it dramatically reduced response reliability. This establishes a clear dissociation between response reliability and response amplitude of the BOLD signal.

Discussion

In this study, we measured neural responses to a real life auditory story and to temporally scrambled and reversed versions of that story. The results revealed a hierarchical topography of TRWs across the human brain in all subjects (Fig. 4). In early auditory regions (A1+), we observed similar responses to the same local elements of the soundtrack, regardless of the coherence of the temporal structure (whether played backward or scrambled at different temporal resolutions), suggesting that responses in these areas are driven primarily by the momentary features of the auditory input. In areas with an intermediate TRW, the level of coherent structure required for reliable responses coincided with the time scale of single sentences, regardless of whether the individual sentences were placed within a longer meaningful structure. Finally, we found areas with a very long temporal window at the upper end of the hierarchy. These areas responded reliably only in cases where sentences were embedded within coherent paragraphs and, in some cases, only when the paragraphs were organized to form a meaningful story.

The topography of TRWs parallels the topography of SRFs. For example, when still images of faces and cars were broken into increasing numbers of pieces and scrambled, a similar hierarchical axis was found across space (Lerner et al., 2001). That is, a gradual transition was observed from sensitivity to spatially local object features (in early visual areas) toward more global and holistic representation (in higher-order visual areas).

Time scale of processing and functional specialization

Although knowledge of the size of the TRW does not determine the function of a brain area, the TRW does constrain the possible functional specializations of an area. The idea that networks of brain regions can provide context for one another's processing has been proposed previously (McIntosh, 2000). However, for reasons of experimental control in the subtraction paradigm, previous imaging studies have mostly evaluated the processing of either short semantic units [e.g., processing of isolated words without context (Booth et al., 2002; Marinković, 2004)] or sentences (Hashimoto and Sakai, 2002; Homae et al., 2002; Friederici et al., 2003). Studies that have evaluated contextual processing at several levels simultaneously within the same experimental design (Mazoyer et al., 1993; Xu et al., 2005) have focused on the classic linguistic distinctions between phonological, lexical, or semantic processing.

In this study, parametric manipulation of the temporal structure revealed three distinct processing stages of the incoming audio soundtrack. First, perceptual sensory areas (Fig. 3, red) analyze low-level acoustic properties and encode low-level linguistic information, such as formant transitions in stop consonants (~ 20 –40 ms) and single syllables (~ 150 –300 ms) (Poeppel, 2003). These areas are optimized for rapid processing of the instantaneous audio properties of a stimulus, regardless of its content or meaning, and they respond similarly to the intact and the time-reversed (meaningless) playback of the same story. Second, higher-order linguistic areas process lexical items and grammatical structures within the boundaries of single sentences (Fig. 3, yellow and green). Such a hierarchical division is consis-

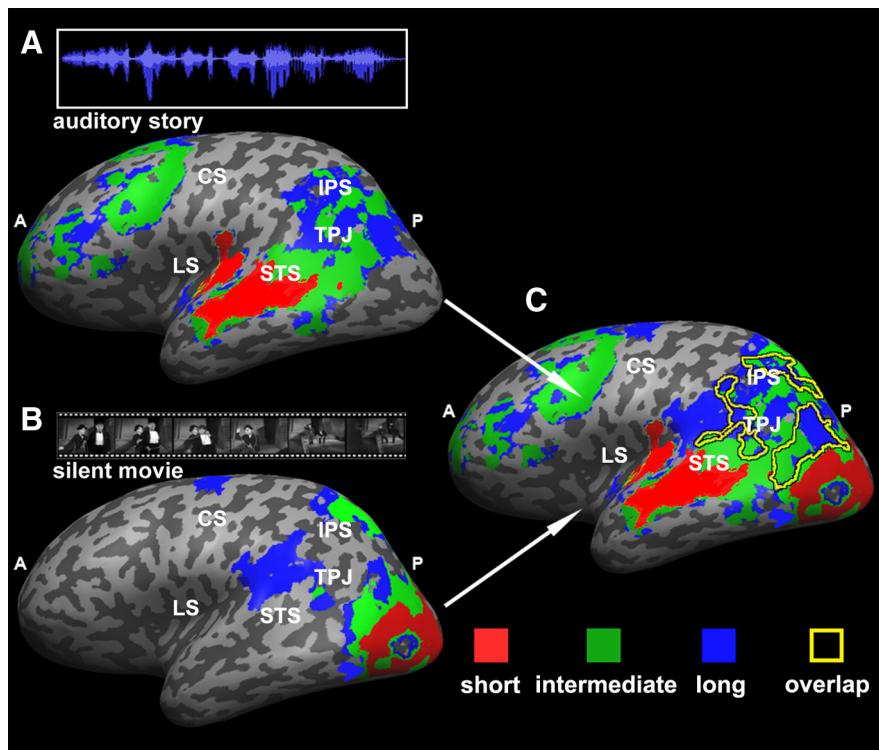


Figure 8. An integrated view of the proposed hierarchy of time scales in visual and auditory modalities. **A**, The hierarchical temporal topography in the auditory domain. **B**, The hierarchical temporal topography in the visual domain [adapted from Hasson et al. (2008)]. **C**, Overlay of the results from **A** and **B**. The yellow line indicates the overlap, which occurs mostly in BA 39 and BA 40, and partially in lateral BA 7 and posterior BA 22. A, anterior; P, posterior.

tent with the observation in a recent study that core auditory regions exhibit high levels of sensitivity to acoustic features, whereas downstream auditory regions in the pSTS show greater sensitivity to speech intelligibility (Okada et al., 2010). Third, extra-linguistic areas (Fig. 3, blue) extract the meaning and context embedded in both the intact paragraphs and the story as a whole. This extended network of areas is involved in comprehension of the fully intact narrative. The extra-linguistic areas include the precuneus, the medial frontal cortex, inferior frontal gyrus, and percingulate cortices. This observation is in agreement with the findings of numerous previous studies examining levels of speech comprehension (Scott et al., 2000; Wilson et al., 2008; Whitney et al., 2009).

Boundaries in a TRW map do not necessarily correspond with boundaries in functional specialization. Consider, for example, that the memory trace of a previous sentence can provide contextual pragmatic cues for the processing of sentences that follow (Carpenter et al., 1995). Considering this, the processing time scale of a sentence-processing area should be larger than the time scale of single sentences. In that respect, it is noteworthy that posterior areas within the supramarginal gyrus exhibit sensitivity to the paragraph time scale, although they are part of the classically defined linguistic network. Moreover, we did not observe sharp boundaries between different levels of processing, but rather a gradual transition from one stage to the next (see supplemental Fig. 1, available at www.jneurosci.org as supplemental material).

TRWs in relation to working memory and predictive coding

Our hypothesis, that each brain area accumulates information over its preferred time scale, is a novel one. As a long-term goal,

we hope to link the notion of TRWs with the influential bodies of research on working memory (Miller, 1960) and predictive coding (Rao and Ballard, 1999; Kiebel et al., 2008). Working memory (WM) is the capacity for actively maintaining information from the recent past to process and act in the present (Baddeley, 1999; Miyake, 1999). Although existing models of WM differ, they focus on how attentional or executive processes help to maintain representations of information in the face of interference or distraction (Baddeley, 1999; Engle et al., 1999; Cowan, 2001; Jonides et al., 2008). Standard WM models can account for data from controlled experiments in which information is maintained over time, but they make few predictions about how the brain accumulates complex temporal structures. The TRW framework is an attempt to characterize how information is actively integrated (as opposed to statically maintained) over time. The results suggest that the temporal integration of information is a distributed and hierarchical process. It is distributed in that each brain area has intrinsic capacity to process natural stimuli over time. It is hierarchical in that the capacity of each TRW increases from early sensory areas to higher-order

perceptual and cognitive areas. Furthermore, whereas most WM paradigms try to account for memory in artificial and abstract situations (e.g., remembering lists of unrelated items), our paradigm aims to study more naturalistic situations in which the incoming information has to be continuously accumulated and integrated with prior context. This notion is more closely related to the notion of long-term working memory suggested by Ericsson and Kintsch (1995), in which, after extensive experience in a particular domain, the relevant memory representations are altered so that domain-specific memory capacity is dramatically expanded (Van Geuchten and Cheng, 2010).

The TRW notion may also be related to predictive coding. Accumulation of information over time can be used not only for interpreting the incoming input in light of past context, but also for predicting the future. Kiebel et al. (2008) suggested that predictions are signaled throughout the cerebral cortex, and that sensory information is reconciled with an internal model by a hierarchical cascade of corticocortical interactions. Consistent with this proposal, it may be argued that regions with longer TRWs will make predictions about forthcoming events with correspondingly larger time scales. Indeed, in language and communication, predictions have been shown to be essential both for semantic and syntactic processing (Pickering and Garrod, 2007). For example, the beginning of a sentence predicts its closure and the likely content of a sentence is predicted by its narrative context. Future studies will be required for investigating the relationship between the time scales over which neural circuits integrate information from the past and the time scales over which they forecast the future.

Time versus accumulation of information over time

Neurophysiological studies have shown that the spatial receptive field of a neuron can shrink or expand as a function of the task and surrounding information (Moran and Desimone, 1985; Sheinberg and Logothetis, 2001; Furmanski et al., 2004). We believe that similar flexibility should be found for TRWs, given that events of similar content can unfold at varying rates. A sentence spoken at one-third its normal rate does, after all, remain easily intelligible. The size of the TRWs should therefore vary more in accordance with the amount of information conveyed via different semantic units such as words, sentences, and paragraphs than in accordance with any absolute temporal period.

Although information and time scales are intimately connected, they are still distinct aspects of a stimulus. An analogy to the spatial receptive field is, again, illustrative. Just as two areas with the same spatial receptive field size in the inferior temporal cortex can have distinct functional selectivity properties (e.g., one responds to faces and the other to inanimate objects), two areas with similar TRWs can process distinct types of information over time. To demonstrate the generality of the TRW concept, we combined the TRW map obtained using the scrambled sequences of a narrated story (Fig. 3) with the TRW map obtained using the scrambled sequences of a silent movie (Hasson et al., 2008). Superposition of the visual and auditory TRWs maps (Fig. 8) suggests that the orderly hierarchy of TRWs is a general topographic organization principle of the human cortex. The early visual and auditory cortices have short TRWs, whereas the TRWs can be seen to gradually increase as one moves to higher-order areas. Moreover, some of the longer TRWs seem to be multimodal, as they clearly process long temporal structures, whether presented aurally or visually. The maps overlap mostly in high-order areas such as Brodmann areas (BA) 39 and BA 40, which are known to be involved in semantic processing. Some overlap also occurs in the lateral BA 7 and posterior BA 22. Interestingly, the silent movie did not evoke reliable responses in frontal areas. Thus, for this particular dataset, we do not see overlap between the audio and visual stimuli in frontal cortices. However, other movie stimuli do evoke reliable responses in these areas (Jääskeläinen et al., 2008; Hasson et al., 2010), and additional studies will be needed to better understand information accumulation over long time scales and across modalities in frontal cortices.

Conclusion

To conclude, the past is always present; each moment in our life is linked to and bounded by the preceding events. Our study demonstrates that the brain relies on a distributed hierarchical network of brain areas to accumulate information over time. Early sensory cortices, such as primary auditory and visual cortices, can accumulate information over relatively short time scales (up to hundreds of milliseconds), whereas high-order areas can accumulate information over long time scales (up to many minutes). These findings reveal a new topographic organization of TRWs that is analogous to the well established hierarchy of SRF sizes in the visual cortex (Hubel, 1988).

References

- Baddeley AD (1999) Essentials of human memory. Hove, UK: Psychology.
- Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Roy Statist Soc* 57:289–300.
- Benjamini Y, Yekutieli D (2001) The control of the false discovery rate in multiple testing under dependency. *Ann Statist* 29:1165–1188.
- Booth JR, Burman DD, Meyer JR, Gitelman DR, Parrish TB, Mesulam MM (2002) Modality independence of word comprehension. *Hum Brain Mapp* 16:251–261.
- Brainard DG (1997) The Psychophysics Toolbox. *Spat Vis* 10:433–436.
- Carpenter PA, Miyake A, Just MA (1995) Language comprehension: sentence and discourse processing. *Annu Rev Psychol* 46:91–120.
- Cowan N (2001) The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behav Brain Sci* 24:87–114; discussion 114–185.
- DeYoe EA, Carman GJ, Bandettini P, Glickman S, Wieser J, Cox R, Miller D, Neitz J (1996) Mapping striate and extrastriate visual areas in human cerebral cortex. *Proc Natl Acad Sci U S A* 93:2382–2386.
- Engle RW, Tuholski SW, Laughlin JE, Conway AR (1999) Working memory, short-term memory, and general fluid intelligence: a latent-variable approach. *J Exp Psychol Gen* 128:309–331.
- Ericsson KA, Kintsch W (1995) Long-term working memory. *Psychol Rev* 102:211–245.
- Ferstl EE, von Cramon DY (2002) What does the fronto-medial cortex contribute to language processing: coherence or theory of mind? *Neuroimage* 17:1599–1612.
- Ferstl EC, Rinck M, von Cramon DY (2005) Emotional and temporal aspects of situation model processing during text comprehension: an event-related fMRI study. *J Cogn Neurosci* 17:724–739.
- Fletcher PC, Happé F, Frith U, Baker SC, Dolan RJ, Frackowiak RS, Frith CD (1995) Other minds in the brain: a functional imaging study of ‘theory of mind’ in story comprehension. *Cognition* 57:109–128.
- Friederici AD, Rüschemeyer SA, Hahne A, Fiebach CJ (2003) The role of left inferior frontal and superior temporal cortex in sentence comprehension: localizing syntactic and semantic processes. *Cereb Cortex* 13:170–177.
- Furmanski CS, Schluppeck D, Engel SA (2004) Learning strengthens the response of primary visual cortex to simple patterns. *Curr Biol* 14:573–578.
- Golland Y, Bentin S, Gelbard H, Benjamini Y, Heller R, Nir Y, Hasson U, Malach R (2007) Extrinsic and intrinsic systems in the posterior cortex of the human brain revealed during natural sensory stimulation. *Cereb Cortex* 17:766–777.
- Hashimoto R, Sakai KL (2002) Specialization in the left prefrontal cortex for sentence comprehension. *Neuron* 35:589–597.
- Hasson U, Nir Y, Levy I, Fuhrmann G, Malach R (2004) Intersubject synchronization of cortical activity during natural vision. *Science* 303:1634–1640.
- 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.
- Hasson U, Avidan G, Gelbard H, Vallines I, Harel M, Minshew N, Behrmann M (2009) Shared and idiosyncratic cortical activation patterns in autism revealed under continuous real-life viewing conditions. *Autism Res* 2:220–231.
- Hasson U, Malach R, Heeger DJ (2010) Reliability of cortical activity during natural stimulation. *Trends Cogn Sci* 14:40–48.
- Holtzman RE, Familitant ME, Deptula P, Hoyer WJ (1986) Aging and the use of sentential structure to facilitate word recognition. *Exp Aging Res* 12:85–88.
- Homae F, Hashimoto R, Nakajima K, Miyashita Y, Sakai KL (2002) From perception to sentence comprehension: the convergence of auditory and visual information of language in the left inferior frontal cortex. *Neuroimage* 16:883–900.
- Hubel DH (1988) Eye, brain, and vision. New York: Scientific American Library.
- Jääskeläinen PI, Koskenkallio K, Balk MH, Autti T, Kaurämäki J, Pomren C, Sams M (2008) Inter-subject synchronization of prefrontal cortex hemodynamic activity during natural viewing. *Open Neuroimag J* 2:14–19.
- Jonides J, Lewis RL, Nee DE, Lustig CA, Berman MG, Moore KS (2008) The mind and brain of short-term memory. *Annu Rev Psychol* 59:193–224.
- Kiebel SJ, Daunizeau J, Friston KJ (2008) A hierarchy of time-scales and the brain. *PLoS Comput Biol* 4:e1000209.
- Larsson J, Landy MS, Heeger DJ (2006) Orientation-selective adaptation to first-and second-order patterns in human visual cortex. *J Neurophysiol* 95:862–881.
- Lerner Y, Hendler T, Ben-Bashat D, Harel M, Malach R (2001) A hierarchical axis of object processing stages in the human visual cortex. *Cereb Cortex* 11:287–297.
- Malach R, Levy I, Hasson U (2002) The topography of high-order human object areas. *Trends Cogn Sci* 6:176–184.

- Marinković K (2004) Spatiotemporal dynamics of word processing in the human cortex. *Neuroscientist* 10:142–152.
- Mazoyer BM, Tzourio N, Frak V, Syrota A, Murayama N, Levrier O, Salamon G, Dehaene S, Cohen L, Mehler J (1993) The cortical representation of speech. *J Cogn Neurosci* 5:467–479.
- McIntosh AR (2000) Towards a network theory of cognition. *Neural Netw* 13:861–870.
- Miller GA (1960) Plans and the structure of behavior. New York: Holt.
- Miyake A (1999) Models of working memory: mechanisms of active maintenance and executive control. Cambridge: Cambridge UP.
- Moran J, Desimone R (1985) Selective attention gates visual processing in the extrastriate cortex. *Science* 229:782–784.
- Okada K, Rong F, Venezia J, Matchin W, Hsieh IH, Saberi K, Serences JT, Hickok G (2010) Hierarchical organization of human auditory cortex: evidence from acoustic invariance in the response to intelligible speech. *Cereb Cortex* 20:2486–2495.
- Pelli DG (1997) The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat Vis* 10:437–442.
- Pickering MJ, Garrod S (2007) Do people use language production to make predictions during comprehension? *Trends Cogn Sci* 11:105–110.
- Poeppel D (2003) The analysis of speech in different temporal integration windows: cerebral lateralization as ‘asymmetric sampling in time’. *Speech Commun* 41:245–255.
- Rao RP, Ballard DH (1999) Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat Neurosci* 2:79–87.
- Scheich H (1991) Auditory cortex: comparative aspects of maps and plasticity. *Curr Opin Neurobiol* 1:236–247.
- 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.
- Sereno MI, Dale AM, Reppas JB, Kwong KK, Belliveau JW, Brady TJ, Rosen BR, Tootell RB (1995) Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science* 268:889–893.
- Sheinberg DL, Logothetis NK (2001) Noticing familiar objects in real world scenes: the role of temporal cortical neurons in natural vision. *J Neurosci* 21:1340–1350.
- Silver MA, Kastner S (2009) Topographic maps in human frontal and parietal cortex. *Trends Cogn Sci* 13:488–495.
- Stephens GJ, Silbert LJ, Hasson U (2010) Speaker-listener neural coupling underlies successful communication. *Proc Natl Acad Sci U S A* 107:14425–14430.
- Talairach J, Tournoux P (1988) Co-planar stereotaxic atlas of the human brain. New York: Thieme Medical Publishers.
- Telkemeyer S, Rossi S, Koch S, Nierhaus T, Steinbrink J, Poeppel D, Obrig H, Wartenburger I (2009) Sensitivity of newborn auditory cortex to the temporal structure of sounds. *J Neurosci* 29:14726–14733.
- Van Genuchten E, Cheng PCH (2010) Temporal chunk signal reflecting five hierarchical levels in writing sentences. In: *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (Ohlsson S, Catrambone R, eds.), pp 1922–1927. Austin, TX: Cognitive Science Society.
- Wandell BA, Dumoulin SO, Brewer AA (2007) Visual field maps in human cortex. *Neuron* 56:366–383.
- Whitney C, Huber W, Klann J, Weis S, Krach S, Kircher T (2009) Neural correlates of narrative shifts during auditory story comprehension. *Neuroimage* 47:360–366.
- Wilson SM, Molnar-Szakacs I, Iacoboni M (2008) Beyond superior temporal cortex: intersubject correlations in narrative speech comprehension. *Cereb Cortex* 18:230–242.
- Xu J, Kemeny S, Park G, Frattali C, Braun A (2005) Language in context: emergent features of word, sentence, and narrative comprehension. *Neuroimage* 25:1002–1015.