

# Temporal scaling of neural responses to compressed and dilated natural speech

Y. Lerner,<sup>1</sup> C. J. Honey,<sup>1</sup> M. Katkov,<sup>1,2</sup> and U. Hasson<sup>1</sup>

<sup>1</sup>Department of Psychology and the Neuroscience Institute, Princeton University, New Jersey; and <sup>2</sup>Department of Psychology, New York University, New York, New York

Submitted 10 July 2013; accepted in final form 17 March 2014

**AQ: 2** **Lerner Y, Honey CJ, Katkov M, Hasson U.** Temporal scaling of neural responses to compressed and dilated natural speech. *J Neurophysiol* 111: 000–000, 2014. First published March 19, 2014; doi:10.1152/jn.00497.2013.—Different brain areas integrate information over different timescales, and this capacity to accumulate information increases from early sensory areas to higher order perceptual and cognitive areas. It is currently unknown whether the timescale capacity of each brain area is fixed or whether it adaptively rescales depending on the rate at which information arrives from the world. Here, using functional MRI, we measured brain responses to an auditory narrative presented at different rates. We asked whether neural responses to slowed (speeded) versions of the narrative could be compressed (stretched) to match neural responses to the original narrative. Temporal rescaling was observed in early auditory regions (which accumulate information over short timescales) as well as linguistic and extra-linguistic brain areas (which can accumulate information over long timescales). The temporal rescaling phenomenon started to break down for stimuli presented at double speed, and intelligibility was also impaired for these stimuli. These data suggest that 1) the rate of neural information processing can be rescaled according to the rate of incoming information, both in early sensory regions as well as in higher order integrative cortexes, and 2) the rescaling of neural dynamics is confined to a range of rates that match behavioral performance.

fMRI; real-life auditory stimuli; speed of information processing; slow and fast rates of speech

REAL-LIFE EVENTS, SUCH AS listening to a speech or watching a movie, unfold over many minutes. During such events, our brains absorb information continuously for the duration of the experience. The information gathered at each particular moment, however, only becomes meaningful in the context of previous events. For example, the meaning of a word depends on its context within a sentence, and each sentence only achieves full meaning in the context of the larger narrative.

Recently, we suggested that the brain handles this nested temporal complexity by increasing its processing timescale from low level sensory areas to high level frontal and parietal areas (Hasson et al. 2008; Lerner et al. 2011; Honey et al. 2012). By analogy with the notion of a spatial receptive field (SRF), the temporal receptive window (TRW) of a neural circuit can be defined as the length of time before a response during which sensory information may affect that response. TRWs are short in sensory areas and become gradually longer toward higher order areas (see MATERIALS AND METHODS and Fig. 1).

The problem of processing temporally nested information is complicated by the fact that real-life information arrives at varying rates. For example, the fastest American English speakers articulate a sentence about twice as fast at the slowest speakers (Smith 2002). Because human listeners can comprehend spoken language with remarkable robustness to speech rate, their brains must be capable of integrating patterns of information over multiple-timescales (e.g., combining sequences of words within and across sentences), while maintaining a functional invariance to the absolute arrival time of each item.

Numerous brain mechanisms can produce functional invariance to absolute speech timing. For example, rate invariance can be achieved in a system in which the processing timescale is absolutely fixed, so that if speech signals arrive at different rates then a different region is needed to process the same information. In this system, temporal rescaling of the responses would not be observed within any area; instead, the various aspects of speech processing would switch between brain regions as the speech rate is varied. A second model would achieve functional invariance via a specialized memory buffer system (e.g., in the early auditory cortex), which accepts and accumulates information at a variable rate but then transmits its output at a constant (perhaps optimal) rate to higher order brain areas. In such a system we would observe rescaling of the neural responses only up until the buffer (e.g., auditory) level of the neural hierarchy but not in higher order regions. Finally, in a model suggested by Gütig and Sompolinsky (2009), functional invariance can be generically achieved by adaptive scaling of the firing rate of neurons according to the incoming information rate (Gütig and Sompolinsky 2009). The firing rate, and its modulation, are accelerated (decelerated) when the information rate is high (low). Under this model we would expect response rescaling in all regions.

Partial support for the rescaling hypothesis comes from electrophysiological and neuroimaging studies that reported, for early auditory areas, rescaling of neural responses with speech rate (Ahissar et al. 2001; Poldrack et al. 2001; Nourski et al. 2009; Mukamel et al. 2011; Vagharchakian et al. 2012; Peelle et al. 2013). However, rescaling of responses in these areas is unsurprising, given that the responses in early auditory areas, with short TRWs, are locked to the momentary low level properties of the audio envelope, which scales with the speech rate. A crucial test of the rescaling hypothesis concerns the responses in higher order areas, which have longer TRWs. Response rescaling is more challenging in higher order areas, because processing in these areas must be sensitive to the temporal relations over seconds of time (e.g., word sequences

**AQ: 4** Address for reprint requests and other correspondence: U. Hasson, Dept. of Psychology and the Neuroscience Institute, Green Hall, Princeton, NJ 08540-1010 (e-mail: hasson@princeton.edu).

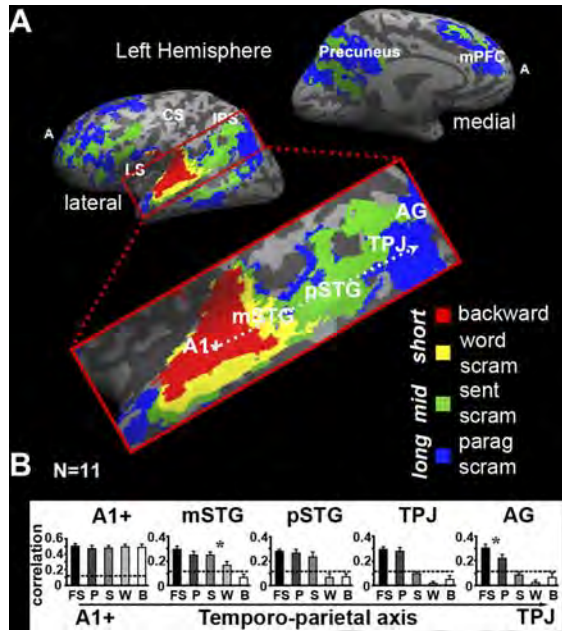


Fig. 1. Identification of regions with short and long temporal receptive windows (TRWs). A: map of TRWs, in which each voxel is colored according to the level of coherent temporal structure that was required to produce significant response reliability [intersubject correlation (inter-SC)] in that voxel. Auditory narratives of varying temporal coherence were presented: the backward stimulus has the least temporal coherence, while the word scramble, sentence scramble, and paragraph scramble stimuli contain increasing levels of temporally coherent information. Voxels are marked according to the least coherent stimulus that could drive reliable responses in that voxel. Thus voxels exhibiting reliable responses to all stimuli (including “backward”) are labeled backward and colored red (see A1+). Voxels that exhibited reliable responses to all stimuli except the backward stimulus are labeled “word scram” and colored yellow [see medial superior temporal gyrus (mSTG)]. Similarly, voxels that were reliable only for the “sentence scramble,” “paragraph scramble,” and the “intact” stimulus are labeled “sentence scram,” and they are colored green [see posterior pSTG (pSTG)]. Finally, voxels were assigned the label “paragraph scram” and colored blue if they responded reliably only to the paragraph scramble stimulus and the intact stimulus [see temporo-parietal junction (TPJ)]. B: reliability profiles for ROIs defined along the A1-TPJ axis. Early auditory areas (A1+) responded reliably to all conditions regardless of the level of temporal scrambling, and thus they have short TRWs. Areas adjacent to A1+ along the superior temporal gyrus exhibited an intermediate TRW. Here, coherent information at the “words” (mSTG), “sentences” (pSTG), or longer timescales was necessary to elicit reliable response time courses. The longest TRWs among the regions of interest (ROIs) were found in the TPJ and angular gyrus (AG). In these regions, reliable responses were evoked only by the paragraph scramble and intact stimuli.

at the start and end of a sentence) while remaining insensitive to absolute time difference (e.g., changes in speech rate).

Here, using functional magnetic resonance imaging (fMRI), we conducted experiments, combined with numerical simulations, to test the prediction that areas with short TRWs and areas with long TRWs can both scale their neural dynamic to match the rate of incoming information. Previous imaging studies reported increased activation for time-compressed speech relative to normal speech in some higher order areas, such as ventral premotor cortex and the posterior superior temporal gyrus (pSTG) (Adank and Devlin 2010; Peelle et al. 2004; Hertrich et al. 2013). However, the increase in response amplitude to compressed speech cannot provide direct evidence for or against the rescaling hypothesis. The rescaling hypothesis predicts that the response time course of the compressed (stretched) speech will be a temporally rescaled ver-

sion of the responses to uncompressed speech, and thus it is independent of changes in the mean response amplitude across conditions.

Directly, testing the temporal scaling hypothesis requires mapping the time-varying response patterns in multiple regions for input of multiple rates. We therefore tested whether the response time course for a slowed (speeded) speech stimulus is a temporally compressed (stretched) version of the response to the original. We presented subjects with slowed versions (150 and 200% duration) as well as speeded versions (50 and 75% duration) of a real-life audio story, while measuring brain responses using fMRI. Subsequently, we rescaled the neural responses to the slowed (speeded) story by compressing (dilating) the neural responses to match the responses to the 100% duration stimulus. Note that the term “neural response” in this article refers to the indirect assessment of the aggregate neuronal responses in each cortical region as captured by the blood-oxygenation level-dependent (BOLD) signal using fMRI. In addition, we quantified the intelligibility of our stimuli at all speech rates, and asked whether temporal scaling is observed even when the auditory narrative is no longer intelligible.

## MATERIALS AND METHODS

### Identifying Regions of Interest with Short and Long Processing Timescales Based on Prior Work

In this study we aimed to test whether the integration size of each TRW scales up (down) with the scaling of the speech rate. To test this, we must first map the TRW hierarchy. To do so, we identified regions with short and long TRWs (Fig. 1A) based on a previously obtained data set (Lerner et al. 2011). Stimuli for the experiment were generated from a 7-min auditory story (“Pie-man”) told by Jim O’Grady at a live storytelling performance (“The Moth” storytelling event, New York City). The TRW of each region was defined according to the response reliability in that region across stimuli scrambled at different timescales: “words” ( $0.7 \pm 0.5$  s), “sentences” ( $7.7 \pm 3.5$  s), and “paragraphs” ( $38.1 \pm 17.6$  s). Early auditory areas with short TRWs (e.g., A1+ in Fig. 1B) responded reliably and similarly at all scrambling levels (i.e., responses were not dependent on temporal history). Further up the processing hierarchy, more and more of the sensory history was found to affect processing in the present moment. This hierarchy was evident both within the anterior portion of the ventral-stream (moving from A1+ toward the temporal pole) and within the dorsal stream [moving from A1+ along the STG toward the angular gyrus (AG); Hickok and Poeppel 2004]. In areas with especially long TRWs, such as the temporo-parietal junction (TPJ) and AG in Fig. 1B, the cortical activity at each moment was reliable only for stimuli that were coherent over long timescales (i.e., those stimuli in which entire paragraphs were kept intact). The definition of regions of interest (ROIs) with short and long processing timescales allowed us to test whether temporal rescaling is observed only in areas that respond primarily to recent stimulus properties [A1+, medial (mSTG)] or also in regions whose responses depend on patterns of information accumulate over seconds or minutes of time (TPJ, AG). For further details, see Lerner et al. (2011).

### Testing for Response Rescaling as a Function of Stimulus Information Rate

In our prior study of temporal integration (Lerner et al. 2011), we maintained a constant information rate while interfering with the temporal ordering of information at different levels of granularity. Here, we preserve the ordering of information while manipulating the

information rate. The rate manipulations enable us to test how the brain adjusts its neural dynamics to changes in the stimulus rate, to achieve a perceptual invariance to speech rate.

**Subjects.** Twenty-four subjects (15 males, ages 18–37 yr) participated in the fMRI study. Twenty additional subjects (11 males, ages 18–29 yr) participated in the behavioral study. All subjects were native English speakers. All subjects were right-handed; had no history of previous neurological, psychiatric, or speech-related symptoms; and reported normal hearing. One subject was removed from the study because of anatomical abnormalities. Three subjects were removed from the analyses due to extensive head motion. In the imaging scans, runs in which head motion was  $>2$  mm or where the signal was corrupted (e.g., contained thermal spikes identified by visual inspection) were discarded from the analysis, and additional subjects were scanned until data from 15 subjects were collected for each condition. Overall, eight subjects participated in all conditions, four subjects in four conditions, three subjects in three conditions, and five subjects in one or two conditions. The procedures were approved by the Princeton University Committee on Activities Involving Human Subjects. All subjects provided written informed consent.

**MRI acquisition.** Subjects were scanned in a 3T head-only MRI scanner (Allegra; Siemens, Erlangen, Germany). A custom radio frequency coil was used for the structural scans (NM-011 transmit head coil; Nova Medical, Wakefield, MA). For fMRI scans, volumes were acquired using a T2\*-weighted echo planar imaging pulse sequence [repetition time (TR) = 1500 ms; echo time (TE) = 30 ms; flip angle =  $75^\circ$ ], each volume comprising 25 slices of 3-mm thickness with a 1-mm gap (in-plane resolution =  $3 \times 3$  mm<sup>2</sup>). Slice acquisition order was interleaved. In addition, a set of 160 T1-weighted high-resolution ( $1 \times 1 \times 1$  mm<sup>3</sup>) anatomical images 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 =  $1 \times 1$

mm<sup>2</sup>, fields of view = 256 mm<sup>2</sup>) and used for three-dimensional (3D) reconstruction. To minimize head movement, subjects' heads were stabilized with foam padding. Stimuli were presented using Psychophysics Toolbox. MRI-compatible headphones (MR Confon; Magdeburg, Germany) were fitted to provide considerable attenuation to the scanner noise and to present the audio stimuli to the subjects.

**Stimuli and experimental design.** FMRI EXPERIMENT. Stimuli for the experiment were generated from the same 7-min auditory story ("Pie-man") that was used to define ROIs based on TRWs (see above). Subjects listened to the intact story at its original rate (100% duration), as well as to temporally speeded or slowed versions of the story (Fig. 2). Specifically, the original story was modified to produce two speeded stimuli (50% duration and 75% duration) and two slowed stimuli (150% duration and 200% duration). To modify the stimulus rate, we used the ChangeTempo algorithm from the open-source SoundTouch audio library (<http://www.surina.net/soundtouch/>) as implemented in Audacity 1.2.6 (<http://audacity.sourceforge.net/>). This algorithm preserved the fine frequency structure of the auditory waveform and sound pitch, while compressing or dilating its energy envelope. Figure 3C presents the distribution of syllables per second for the 100% intact, 50% speeded, and 200% slowed stories. Number of syllables has been counted using <http://www.howmanysyllables.com>. We padded the stimuli with 12 s of music at the beginning, and 3 s of silence at the end. These music and silence periods were discarded from all analyses. A typical session comprised five runs. To minimize the repetition effect, especially in the less intelligible conditions, presentation of all conditions was pseudorandomized across subjects. Attentive listening to the story and its scaled variants was confirmed by brief communication with subjects between runs and using a simple questionnaire at the end of the experiment.

**Behavioral experiment.** To evaluate the intelligibility of each stimulus, behavioral sessions were conducted with a group of subjects that

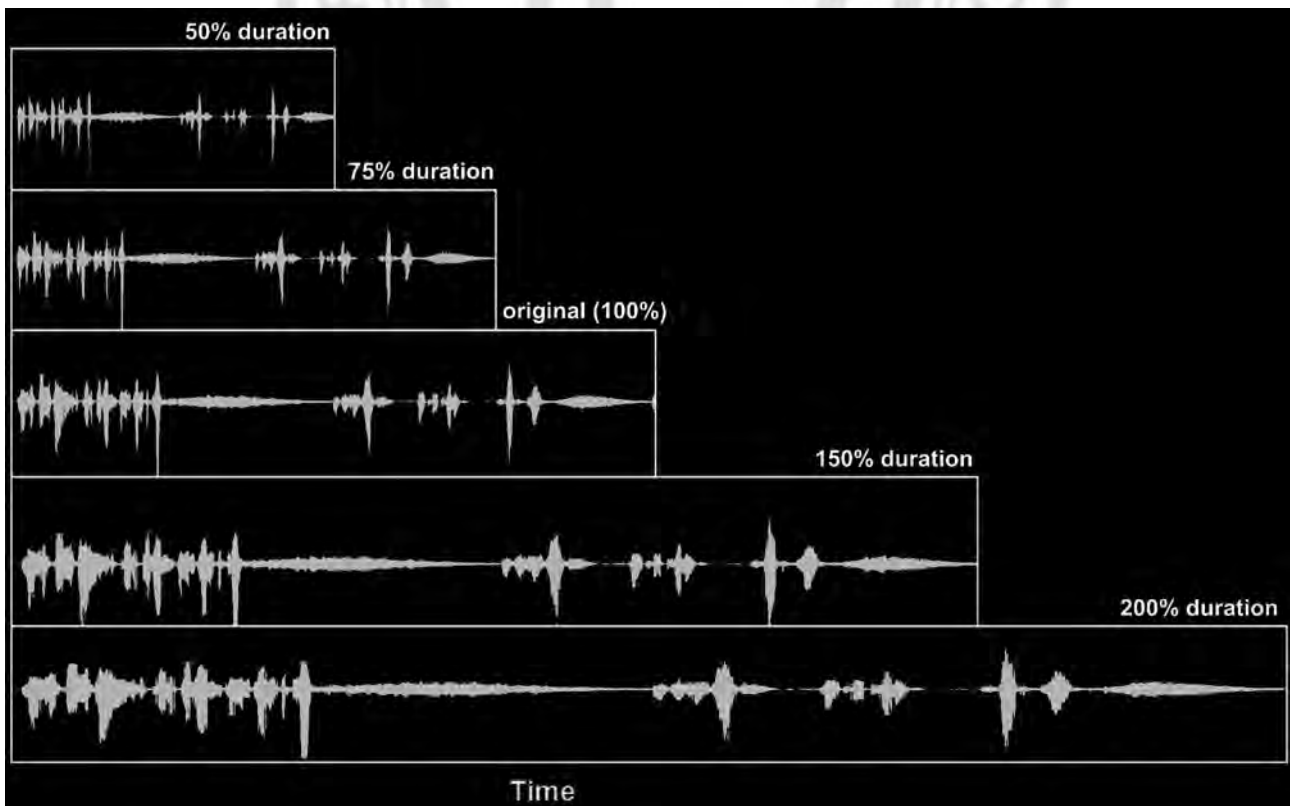


Fig. 2. Examples of stimuli. A schematic representation of the stimuli used in the experiment. The rate of a real life story (100%) was speeded in the time [75% duration and 50% duration (double speeded)] and slowed in time [150% duration and 200% duration (twice as slow)]. To generate speeded and slowed versions of the story, an algorithm was employed that preserved the fine frequency structure of the auditory waveform while compressing or dilating its energy envelope.

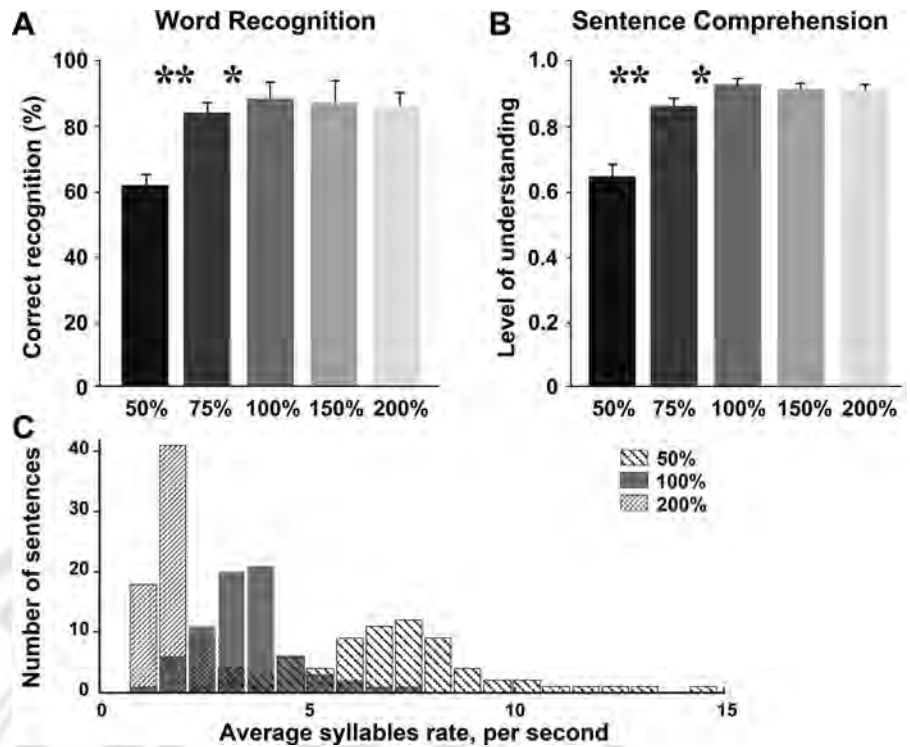


Fig. 3. Behavioral performance. Averaged recognition performance is shown for each condition. *A*: number of correctly recognized words. The y-axis shows the average percentage of correct words recognition across all sentences. *B*: level of sentence comprehension. The y-axis shows the semantic comprehension of a sentence. Error bars indicate  $\pm$  SD. \*\*\*Significance difference between the condition and the intact (100%) story. Note that there was a reduction in the intelligibility of the 50% duration condition across the 2 behavioral estimates. *C*: histogram of syllables rates across sentences for the uncompressed story (100%), the fastest (50%), and the slowest (200%) conditions.

were not involved in the imaging study. To match, as much as possible, the listening conditions in the behavioral experiment to the listening conditions in the MRI scanner, we ran the behavioral tests in a “mock” scanner that simulates the scanner environment (including mimicking of scanner noise). To assess the intelligibility of speech in each condition we modified the stimulus presentation so that each sentence was followed by a silent interval. Subjects were instructed to repeat each sentence immediately after its presentation, during the silent interval. To minimize performance enhancements due to prior exposure with a more intelligible stimulus, each subject heard only a subset of the conditions, with the most speeded (or slowed) condition first. Thus there were two groups: one group ( $n = 10$ ) heard the story at 50% duration, intact (100%), and then 150% duration, while the other group ( $n = 10$ ) heard the story at 200% and then 75% duration. Two measurements of participants’ performance were estimated: 1) number of correctly recognized words, and 2) level of sentence comprehension that evaluated (on a scale from 0 to 1) whether subjects were able to understand the gist of the sentence. Word recognition was operationalized as the mean percentage of words that subjects were able to accurately repeat across all sentences. Sentence comprehension was an estimate of whether subjects could report the gist of the sentence, even in cases when they failed to accurately perceive some of its constituent words. This “gist” sentence comprehension was evaluated by two independent raters who had been familiarized with the story content. The agreement between the raters was high (~95%), and ratings for each condition were averaged across raters and across subjects (Fig. 3, *A* and *B*).

**Numerical simulation.** Numerical stimulation was performed to assess the sensitivity of fMRI in detecting rescaling of the neural response time courses. In particular, we examined the rescaling of the slow, hemodynamic, BOLD signal in cases when the underlying neural responses are precisely accelerated (decelerated) by a known factor. To base our simulation on a realistic estimate of the neuronal response time course, we used as our starting point, intracranial-EEG (iEEG) data previously recorded from the lateral Sylvian fissure, near the early auditory cortex (A1+), in an epileptic patient who listened to the same (100% duration) intact story. The modulations of high frequency (65–150 Hz) power in

this electrode were robustly correlated with the BOLD signal extracted from A1+ in our subjects who listened to the same story (see Fig. 5A). Voltage signals were sampled at 30 kHz using AdTech electrodes as a patient listened to the 100% intact story. Voltage traces were subsequently downsampled to 400 Hz before spectral power modulations were estimated using Slepian tapers (3 tapers, 65–150 Hz band in 1,000-ms time windows overlapping by 900 ms). iEEG power modulations were log-transformed to Gaussianize the signal. The rapid temporal modulations of high-frequency (65–200 Hz) power at this site, when convolved with a hemodynamic response function (HRF), were strongly correlated ( $r = 0.48$ ) with the BOLD signal sampled from the same cortical region in our current BOLD dataset (see Fig. 5A), replicating previous reports linking fluctuations in the BOLD signal and high-frequency power (Mukamel et al. 2005; Privman et al. 2007).

In each run of the simulation we added some noise to the linearly compressed or dilated iEEG signal (see Fig. 5B1) to more closely match the signal-to-noise observed in the empirical data (see Fig. 5B2). The noise was generated by phase-scrambling the original signal (Theiler et al. 1992) and thus had the same power spectrum as the original. The level of noise was adjusted to produce BOLD correlation of similar magnitude to those observed in A1+ for repeated presentations of the story. Next, this altered iEEG signal was convolved with an HRF (see Fig. 5B3) (Glover 1999). Finally, we upsampled (or downsampled) the result of the convolution (see Fig. 5B4) to produce our estimates of what a BOLD response to the speeded and slowed stimuli should look like, under the assumptions of linear temporal scaling of underlying activity. The resulting signal estimates were then cross-correlated with an HRF-convolved version of the original iEEG signal (see Fig. 5C).

**data analysis.** PREPROCESSING. fMRI data were analyzed with the BrainVoyager QX software package (Brain Innovation, Maastricht, The Netherlands) and with in-house software written in MATLAB (MathWorks, Natick, MA). Preprocessing of functional scans included intrasession 3D motion correction, slice scan time correction, linear trend removal, and high-pass filtering (3 cycles per experiment). Spatial smoothing was applied using a Gaussian spatial filter of 6-mm full-width at half-maximum value to correct

for misalignments across subjects and structural differences between brains. 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 main outcomes of the study were derived from the intersubject correlation (inter-SC) analysis. This analysis provides a measure of the reliability of the responses to natural temporally complex stimuli by assessing the similarity of the BOLD response time courses across different subjects (Hasson et al. 2010).

Correlation maps were constructed on a voxel-by-voxel basis (in Talairach space) within each condition (Fig. 4) and between conditions (see Fig. 7) by comparing the responses across all subjects in the following manner. For each voxel, the Pearson product-moment correlation  $r_j = \text{corr}(TC_j, \overline{TC}_{All-j})$  between that voxel's BOLD time course  $TC_j$  in one individual and the average  $\overline{TC}_{All-j}$  of that voxel's BOLD time courses in the remaining individuals was computed. The average correlation  $R = 1/n \sum_{j=1}^n r_j$  was then calculated at every voxel.

**CORRELATION BETWEEN CONDITIONS.** Correlation maps between conditions (see Fig. 7) and ROI analysis (see below and Fig. 8) were computed relative to an independent group of 11 subjects who listened to the intact (100%) condition in a separate prior study (Lerner et al. 2011). First, time courses in the speeded and slowed conditions were linearly interpolated (using the "interp1" Matlab function) to have the same number of time samples in all conditions. Next, for every subject from the independent group ( $n = 11$ ), the time course was correlated with the averaged resampled time course of 50, 75, 100, 150, and 200% duration conditions in the current study. Then the correlations were averaged across subjects.

**PHASE-RANDOMIZED BOOTSTRAPPING.** Statistical significance of each observed correlation was assessed using a bootstrapping procedure based on phase randomization. Phase-randomization was performed by fast Fourier transforming the signal, randomizing the phase of each Fourier component, and then inverting the Fourier transformation. This procedure leaves the power spectrum of the signal intact but removes temporal alignment of the signals. For every empirical time course in every voxel, 5,000 bootstrap time series were generated using a phase-randomization procedure. With the use of these bootstrap time courses, a null distribution of the correlation values was determined for each subject in each voxel, and these were combined to produce a null distribution of the average correlation. The  $P$  values of the empirical average correlations were computed by comparison with these null distributions.

**FALSE DISCOVERY RATE CORRECTION.** Finally, to correct for multiple comparisons, the Benjamini-Hochberg-Yekutieli false-discovery procedure that controls the false discovery rate under assumptions of dependence was applied (Benjamini and Hochberg 1995; Benjamini and Yekutieli 2001). Following the procedure,  $P$  values were sorted in ascending order and the value  $p_{q^*}$  was chosen. This value which is the  $P$  value corresponding to the maximum  $k$  such that  $p_k < k/nq^*$ , where  $q^* = 0.05$  is the false discovery rate threshold, and  $n$  is the total number of voxels.

**RESULTS**

Speech perception is robust across a wide range of speech rates. In this study we asked whether the perceptual invariance is achieved by a dynamical scaling (speeding/slowing) of the underlying processes in response to changes (speeding/slowing) in speech rate. To test this prediction we asked whether regions with long integration windows rescaled their responses

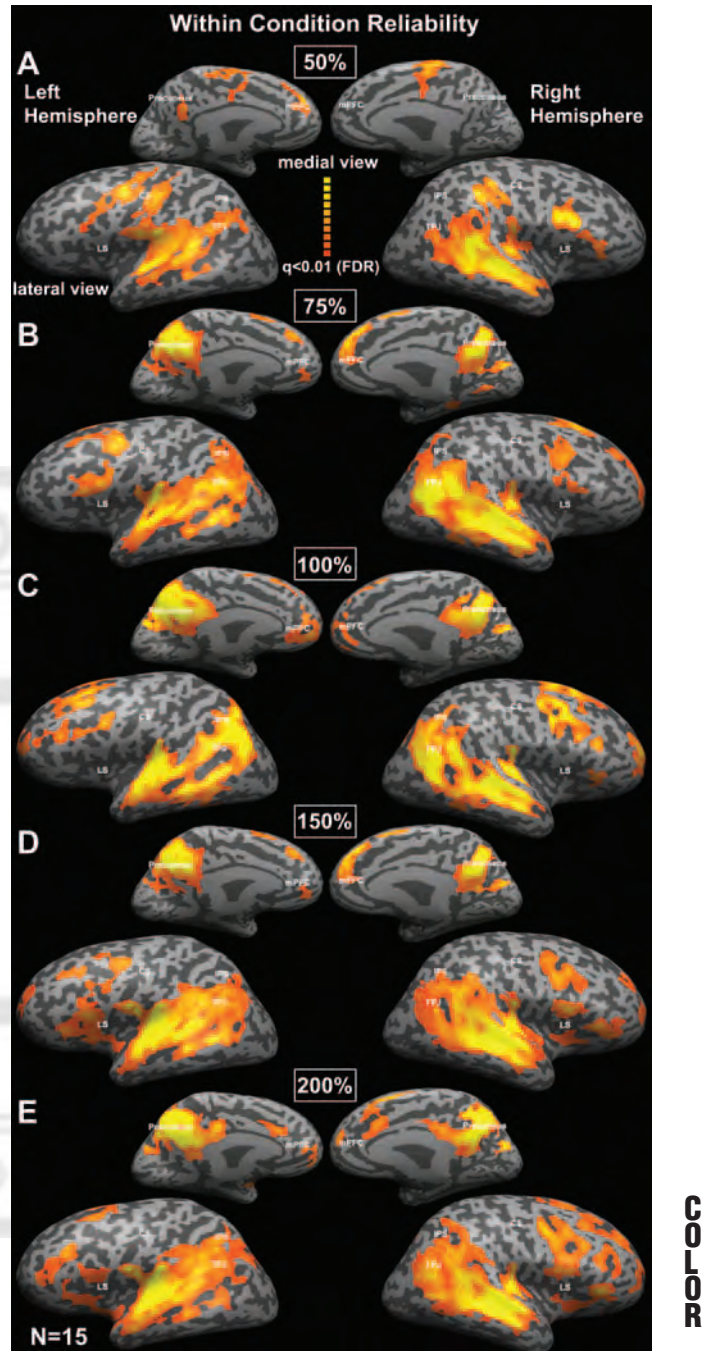


Fig. 4. Reliability of responses for each presentation rate. Maps of the reliability of responses across subjects ( $n = 15$ ) were computed separately for each stimulus condition and superimposed on an inflated brain shown in lateral and medial views. The maps illustrate the extent of the inter-SC for the 50% duration (A), 75% duration (B), intact story (C; 100%), 150% duration (D), and 200% duration (E). LS, lateral sulcus; CS, central sulcus; IPS, intraparietal sulcus; mPFC, medial prefrontal cortex.

according to the information rate of the stimulus. To that end, we presented a spoken auditory story in an intact form (100% duration), as well as in two temporally speeded (75% duration and 50% duration) and two temporally slowed forms (150% duration and 200% duration). The stimulus waveforms are schematically presented in Fig. 2.

In the first section of the results, we report behavioral recognition performance for each speech stimulus, as well as

F4

ROFOL

the set of brain regions that respond reliably at each stimulus rate. In the second section, we test the prediction that neural response time courses temporally rescale according to the rate of incoming information.

### *Recognition Performance and Neural Response Reliability as a Function of Rate*

**Behavioral recognition performance.** Speeding and slowing down the story may reduce the intelligibility of the story. Thus, to estimate the effect of speeding and slowing on the comprehensibility of a story, we assessed subjects' word recognition and sentence comprehension at each level of scaling in a "mock" scanner that mimics the scanner noise. Word recognition was significantly reduced for both speeded conditions [one-tailed paired  $t$ -test, 50% duration:  $t(9) = 54.7$ ,  $P < 5.7 \times 10^{-13}$ ; and 75% duration:  $t(9) = 2.8$ ,  $P < 0.01$ ]. Sentence comprehension results indicated that slowing the story did not reduce the overall level of intelligibility of either of the slowed conditions, relative to the intelligibility of the 100% duration ( $94 \pm 1\%$ ). However, speeding the story reduced the intelligibility of the story to  $87\% \pm 3\%$  for the 75% duration condition [ $t(9) = 5.3$ ,  $P < 0.0002$ , one-tailed paired  $t$ -test] and to  $65\% \pm 3\%$  for the 50% duration condition [ $t(9) = 33.2$ ,  $P < 4.93 \times 10^{-11}$ , one-tailed paired  $t$ -test; Fig. 3, *A* and *B*]. The reduction in the 50% condition may be related to the scanner noise or to the quality of the recording. Also, it may be related to the speaker's relatively fast speech rate, reported in Fig. 3*B*. However, even in the double speed (50% duration) condition, average word recognition and sentence comprehension levels remained  $>60\%$ . Namely, in word recognition:  $62 \pm 3\%$  in 50% condition,  $84 \pm 3\%$  in 75% condition,  $89 \pm 5\%$  in 100% condition,  $87 \pm 7\%$  in 150% condition, and  $87 \pm 4\%$  in 200% condition. In sentence comprehension, it was  $65 \pm 3\%$ ,  $87 \pm 2\%$ ,  $94 \pm 1\%$ ,  $92 \pm 2\%$ , and  $91 \pm 2\%$ , correspondingly.

**Neural response reliability for speech of different rates.** The reliability of responses for each stimulus rate was assessed using inter-SC analysis (Hasson et al. 2004). Figure 4 shows the inter-SC maps (*A–E*) for each of the conditions: 50, 75, 100, 150, and 200% durations. All stimuli (intact, speeded and slowed speech) evoked reliable responses across a widespread cortical network starting with the primary auditory cortex and nearby tonotopic areas (A1+) (Romanski and Averbek 2009) and extending to the linguistic and extra-linguistic regions. Specifically, reliable responses have been found in the STG, AG, supramarginal gyrus, TPJ, parietal lobule, and inferior frontal gyrus, which are known to be involved in various aspects of linguistic processing (Friederici 2002, 2012; Hickok and Poeppel 2007; Rauschecker and Scott 2009) and in the precuneus, inferior occipital gyrus, dorsolateral prefrontal, orbitofrontal, and medial prefrontal cortices (mPFC), which seem to be associated with processing the narrative and content of the story (Xu et al. 2005; Lerner et al. 2011).

The response reliability maps were in agreement with the behavioral findings, as we observed a slight reduction in the reliability for the double speed (50% duration) condition. The reduction was most apparent in the precuneus and mPFC. Thus, when the information stream was fully intelligible, responses were highly reliable across the entire network of brain areas that process the story, regardless of changes in absolute rate. However, when the level of intelligibility decreased, the reliability of responses also decreased.

We next attempted to match the responses for the speeded (slowed) recordings to the intact (100%) responses by upsampling (downsampling) the neural signal. This rescaling analysis enabled us to directly test the prediction that neural responses in areas with short and long TRW are linearly scaled in time (speeded or slowed) in accordance with the rate of incoming speech information. Before conducting the rescaling analysis, we ran a numerical simulation to demonstrate that, despite its low temporal resolution and hemodynamic mediation, fMRI can nonetheless be used to test our hypothesis of linear response rescaling.

### *Temporal Scaling of Neural Responses with Speech Rate*

**Numerical simulation to validate rescaling procedure.** Given the low temporal resolution of the BOLD signal, can fMRI be used to measure the rescaling of neural responses hypothesis? In other words, even if the underlying neural responses are scaled linearly in the speeded or slowed conditions, will the hemodynamics of the BOLD signal interfere with our ability to detect such temporal scaling in the empirical fMRI data? To assess this, we used an iEEG-based numerical simulation to characterize the effect of hemodynamic temporal blurring on the basic comparability of the upsampling (downsampling) BOLD responses and the intact (100%) responses (see MATERIALS AND METHODS and Fig. 5 for details).

Simulation results indicated that, in cases of actual rescaling of the underlying neural dynamics, the process of upsampling and downsampling of the BOLD responses successfully recovers the scaled underlying neural responses signals, despite the presence of the HRF (Fig. 5*B*). However, the interpolated signals must be shifted forward or backward to varying degrees for each condition (varying cross-correlation peaks, Fig. 5*C*) to temporally align with the BOLD response time course for the intact (100%) condition.

**Empirical observations of temporal scaling with speech rate.** Encouraged by the reliability of the BOLD responses within conditions (Fig. 4) and by the simulation (Fig. 5), we next tested whether the underlying neural responses within each brain area are indeed linearly scaled with changes in the absolute rate of presentation. Although we analyzed data throughout the brain, for illustration of the analysis procedure, we first present the average mean time courses from the TPJ (Fig. 6). The region was defined using an independent data set (Lerner et al. 2011), in which we established that the TPJ has a relatively long temporal integration window ( $>30$  s, see Fig. 1). Figure 6*A* shows the time courses in the TPJ averaged across participants in each of the five conditions; Fig. 6*B* shows the responses in the speeded and slowed conditions after upsampling and downsampling, respectively. Each condition time course was shifted to correct for temporal misalignments due to the filtering of the neuronal responses with a fixed HRF (see simulation in Fig. 5). The 75% upsampled and the 150 and 200% downsampled time courses were highly correlated with the intact (100%) response time courses indicating that responses in the TPJ scale up or down linearly in response to temporal scaling of the external speech. However, the correlation of the 50% upsampled time course and the intact signal was weaker. To evaluate the statistical significance of these results, we computed the correlation between the 100% time course and each condition in each of our subjects (see MATERIALS AND METHODS) and compared the reliability across condi-

F5

F6

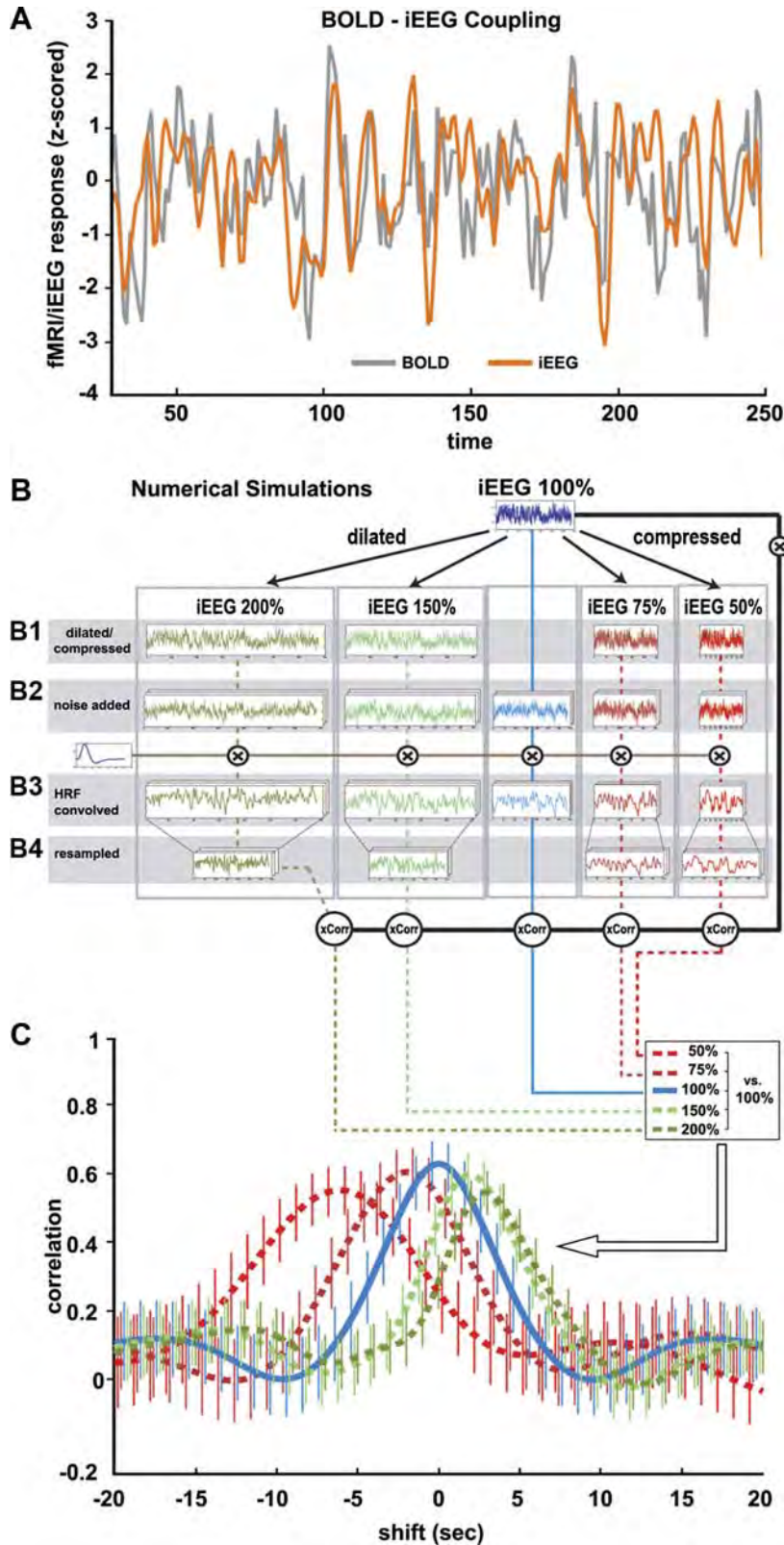
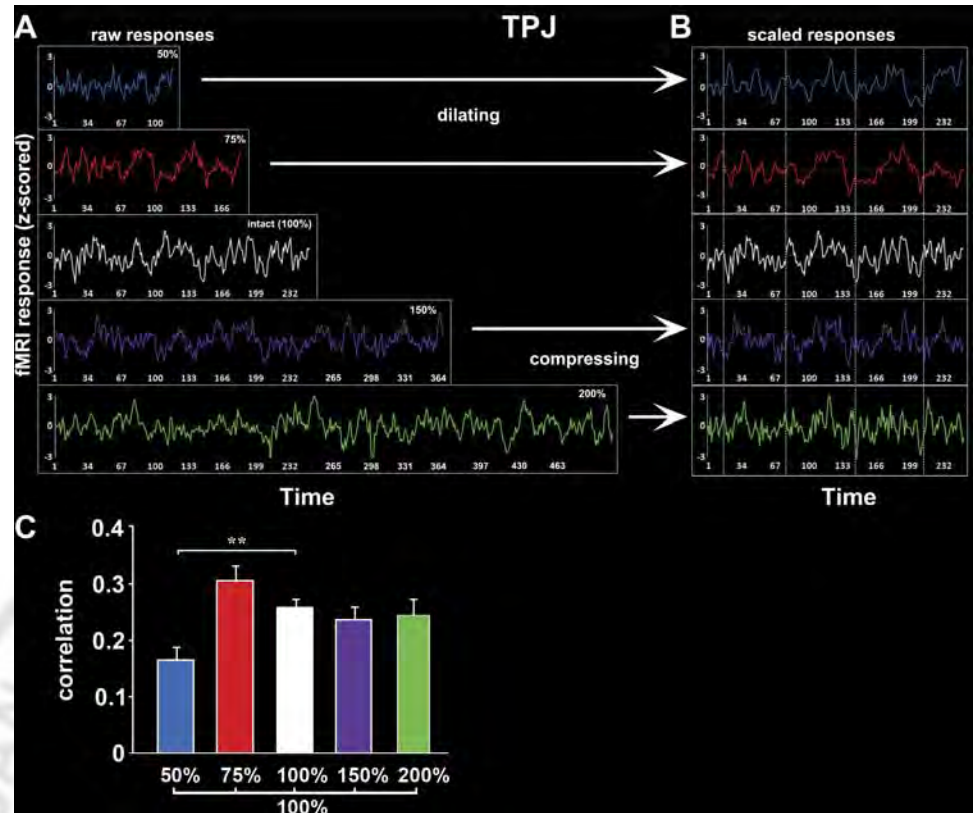


Fig. 5. Simulation pipeline and results. *A*: simulation is founded on strong agreement between 1) the downsampled and convolved intracranial-EEG (iEEG) power recorded from an electrode at the lateral Sylvian fissure in an individual epileptic patient listening to the 100% condition, and 2) averaged blood-oxygenation level-dependent (BOLD) responses sampled from the same area in healthy individuals hearing the 100% condition. The strong agreement between the signals is striking given that the BOLD and iEEG signals were acquired from different subjects and different experimental setups. FMRI, functional MRI. *B*: modeling steps: the electrophysiological response was downsampled (or upsampled) to mimic a linear scaling of the neural signal in response to differing stimuli (*B1*). Spectrum-matched noise was added to the signals to produce simulated BOLD correlation values which match those obtained empirically (*B2*). Noisy rescaled iEEG signals were then convolved with an a hemodynamic response function (HRF; *B3*). Next, the compressed (or dilated) BOLD response time courses were upsampled (or downsampled) to the same time-base as the BOLD responses to the 100% duration story (*B4*). *C*: cross-correlation of the HRF-convolved speeded (slowed) iEEG signals and the HRF convolved original iEEG signal, after up (down) sampling. The results of the simulation procedure indicate that in cases where the underlying neural responses are linearly scaled, the upsampled and downsampled BOLD signals will be highly correlated with the convolved original signal but with a phase lag.

tions (Fig. 6C). First, the responses in the 75% upsampled time course, 150% downsampled time course, and 200% downsampled time course were highly correlated with the intact 100% time course ( $r = 0.32 \pm 0.03$ ,  $r = 0.23 \pm 0.03$ , and  $r = 0.24 \pm 0.03$ , respectively). Second, the reliability of the res-

caled signal was as high as the reliability within the intact 100% condition [one-tailed paired  $t$ -test,  $t(14) = 1.76$ ,  $P = 0.05$  in 75% condition;  $t(14) = 0.79$ ,  $P = 0.22$  in 150% condition; and  $t(14) = 0.34$ ,  $P = 0.37$  in 200% condition]. Third, only for the 50% upsampled time course the correlation

Fig. 6. Response reliability in TPJ for different presentation rates. **A**: raw responses in the TPJ averaged across all participants in each of the five conditions. **B**: dilated of the speeded 50 and 75% duration responses; and compressed of the 150 and 200% duration responses. Note that after accounting for linear response scaling, the responses to the compressed, dilated, and intact stimuli were highly correlated within TPJ. **C**: correlation of the individual time courses to the intact (100%) story in an independent group of listeners ( $n = 11$ ) and the averaged time courses to the speeded, slowed, and intact conditions ( $n = 15$ ).



with the intact 100% time course was significantly lower than the reliability within the 100% condition [ $t(14) = 2.98, P < 0.005$ , one-tailed paired  $t$ -test].

The same analysis was repeated across the whole brain by computing, for each voxel, the correlation between the rescaled (upsampled or downsampled) responses and the responses to the intact (100%) story (Fig. 7). In the double speed (50% duration) condition, the responses deviated considerably from the responses seen for the intact story across most areas (Fig. 7A). However, we observed high correlations between the responses to the intact story and the upsampled 75% duration (Fig. 7B) as well as the downsampled 150 and 200% durations (Fig. 7, D and E) time courses, across all regions that responded reliably to the intact story. Moreover, the correlation across conditions was as strong as the reliability within the 100% (unaltered) condition (Fig. 7C).

An ROI analysis (see MATERIALS AND METHODS) revealed similar results in A1+, pSTG, TPJ, and precuneus (Fig. 8A). Here, the individual responses to the intact (100%) story were correlated with the averaged responses (across subjects) to the speeded, slowed, and 100% conditions. Although our simulation indicates that we should be able to detect a linear scaling of responses in the double speeded (50% duration) condition, the 50% duration stimulus showed a reduced match with the 100% condition. This reduction corresponds with a reduction in behavioral performance for the 50% rate.

## DISCUSSION

Human listeners comprehend spoken language with a remarkable robustness to changes in speech rate (Smith 2002). Behavioral studies, using time-compressed speech, have shown

that speech can still be understood even when played at twice or half of the original speed (Foulke 1971; Chodorow 1979; Wingfield 1975, 1996; Summerfield 1981; Miller et al. 1984; Wingfield et al. 2003). What is the neural mechanism by which the brain becomes so resilient to variations in speech rate? Our results suggested that the brain achieves a functional invariance to speech rate by temporally scaling (i.e., slowing or speeding) the neural responses to match the speech rate. Such neural scaling was observed at all stages of the processing stream, ranging from early auditory areas to linguistic and extra-linguistic areas. These results imply that the human brain's mechanisms for processing temporally extended information do not work on absolute timescales. Instead, temporal accumulation occurs on a relative timescale, and both short timescale and long timescale regions can flexibly rescale their processing according to the rate at which they receive information. This flexibility of integrative processing suggests that TRW size should be measured in informational units. In addition, the temporal rescaling of neural responses may underlie the robustness of human perception and cognition to variations in absolute timing in an ever-changing world (Gütig and Sompolinsky 2009).

Temporal scaling of responses was an expected outcome in early auditory areas (e.g., A1+, Fig. 8B), which respond mostly to transient stimulus features. An area with short processing timescale that responds primarily to the instantaneous amplitude envelope of auditory input should scale its responses linearly with the scaling of the stimulus envelope. Indeed, prior studies reported such temporal scaling in early sensory areas (Ahissar et al. 2001; Poldrack et al. 2001; Luo and Poeppel 2007; Obleser et al. 2008; Nourski et al. 2009; Mukamel et al. 2011; Vagharchakian et al. 2012). By contrast,



temporal scaling in higher order regions is not guaranteed because in these regions the momentary responses result from integrating information over seconds of past time. Integration of temporally varying information over hundreds of milliseconds poses a difficult computational challenge to the brain (Gütig and Sompolinsky 2009).

The slowing and speeding manipulation allows us to keep the total amount of information delivered by the story constant (as indicated by the intelligibility of the speech across conditions, Fig. 3), while varying the rate of information flow. In contrast, in the scrambling manipulation (Lerner et al. 2011) we kept the temporal rate constant, while disrupting the temporal ordering of information at different granularities. Taken together, our temporal scaling results (Fig. 8A) and temporal scrambling results (Fig. 8B) suggest that each brain area integrates information on a preferred timescale that corresponds to a certain kind of information. For example, a given area's preferred timescale may coincide with the timescale of single words, while another brain area's preferred timescale may coincide with the timescale of single sentences (Lerner et al. 2011). However, because of the natural variability in input rate (due to the variable length of words and sentences and variability in speech rate), the preferred timescale at each level of the processing stream is not fixed but rather operates within a range over which the relevant information typically arrives. Thus, just as SRFs can rescale their receptive field size as a function of task, context, and attention demands (Zipser et al. 1996; Op De Beeck and Vogels 2000; Sheinberg and Logothetis 2001; Li et al. 2006), the TRWs of brain regions exhibit flexibility in their duration. In particular, the TRWs vary according to the rate at which information is arriving.

There are limits to the temporal flexibility of information processing. This is apparent in the fact that temporal rescaling of responses was not successful across all rates. Although responses were linearly scaled in response to stimuli of 75–200% duration, the response scaling was reduced for the double speed (50% duration) condition, which is also a condition where a reduction in behaviorally assessed intelligibility was observed (Foulke and Sticht 1969; Beasley et al. 1980; Ghitza and Greenberg 2009). This was observed in most brain areas (beside pSTG), including early auditory areas, as A1+, and higher order areas, as the TPJ and precuneus. Thus the flexibility of TRWs is likely to reflect the range of timescales on which information arrives in ecological settings. Moreover, the responses in A1+ were less correlated with both the rescaled stretched and rescaled compressed responses, relative to the uncompressed version (100%). One explanation for this effect is that A1+ is especially tuned to the fine spectrotemporal features that become distorted during the manipulation of the speech rate. In contrast, higher order regions are more accommodating to temporal variability in their input and show equivalent reliability across the uncompressed and rescaled responses. We note that, although comprehension of 200% duration stimulus was intact in our behavioral assay, this condition may nonetheless have placed increasing demands on working memory. Behavioral and neural correlates of this

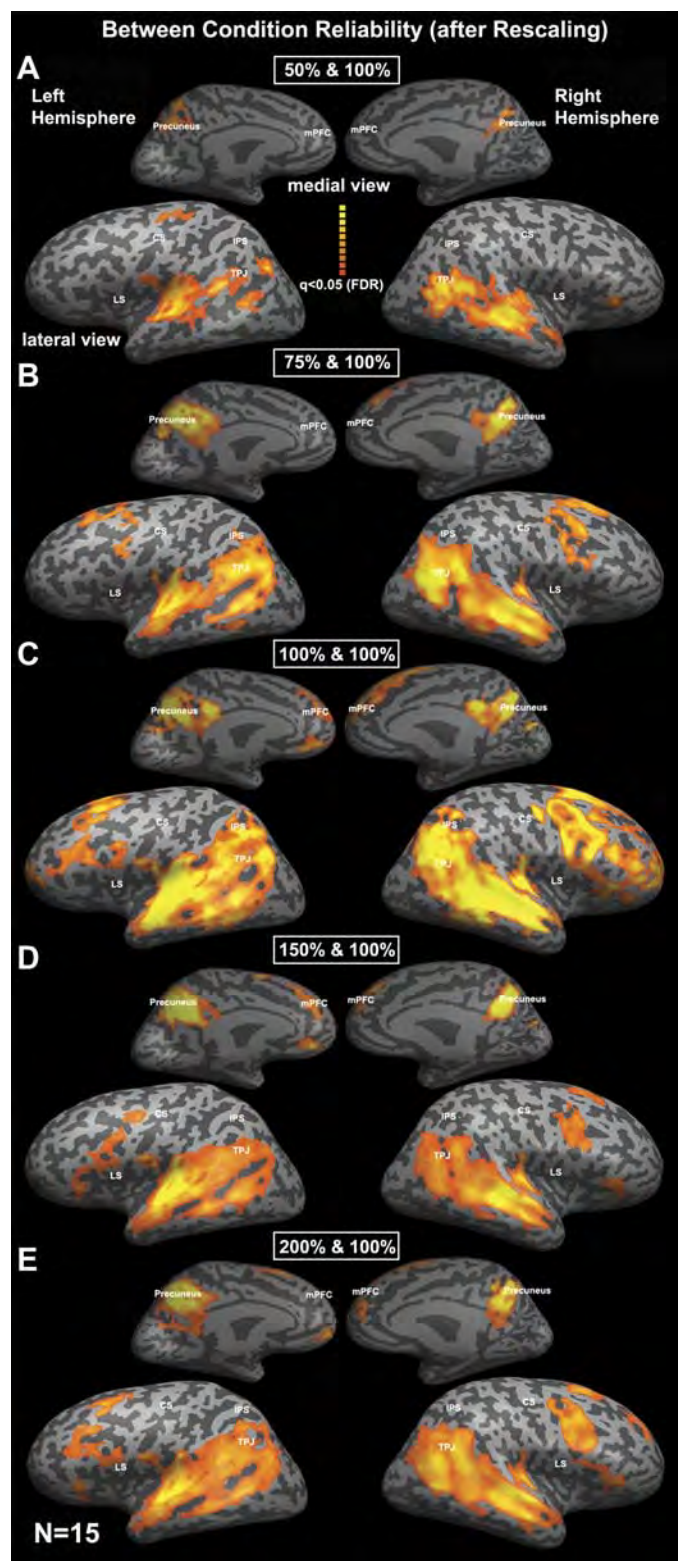


Fig. 7. Topography of rate modification effects across conditions. Maps of the reliability of responses were computed by comparing responses in the scaled speeded and slowed conditions with the responses to the intact story in the previous study (Lerner et al. 2011). The extent of the inter-SC is presented for 2 slowed speeded conditions: 50% duration (A) and 75% duration (B), and 2 compressed slowed conditions: 150% duration (D) and 200% duration (E). C: illustrates the inter-SC between responses to the intact 100% story in the previous and current studies. Note that the scaled responses in the slowed and speeded conditions were similar to the responses in the intact condition in all areas and all conditions beside the 50% duration condition in which we observed a weaker agreement. The figure layout is identical to that in Fig. 4.

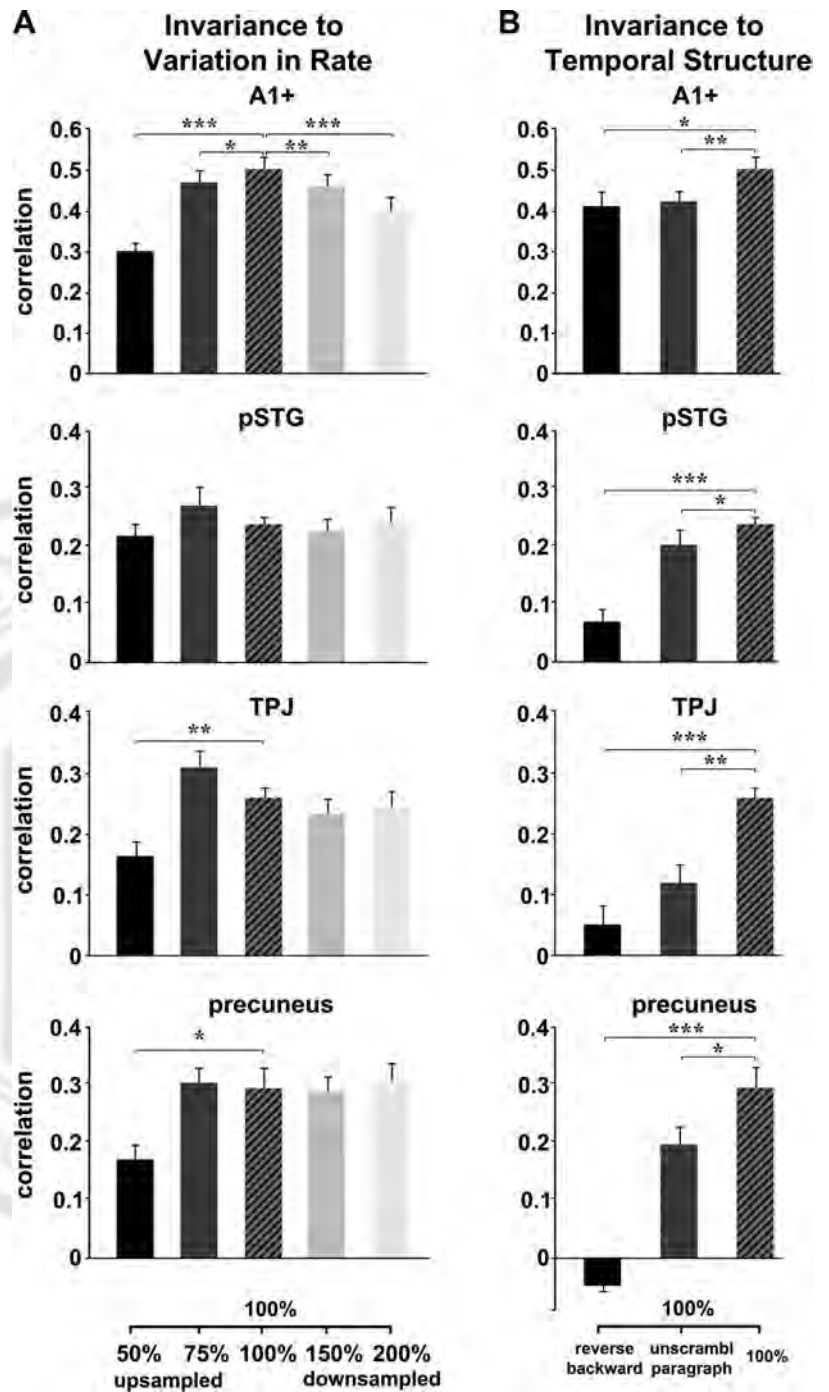


Fig. 8. Response reliability profiles from different ROIs. A: correlation of the individual responses to the intact (100%) story in an independent group of listeners ( $n = 11$ ) and the averaged responses to the speeded, slowed, and intact conditions ( $n = 15$ ) are computed. Early auditory areas exhibited high reliability in all conditions, regardless of the speed modifications. pSTG exhibited intermediate sensitivity to timing. Finally, the strongest sensitivity to stimulus rate was observed in higher order areas, such as TPJ and precuneus. B: correlation of the individual responses to the intact (100%) story and averaged unscrambled conditions (“reverse backward,” “unscrambled paragraphs,” and “unParag”) in Lerner et al. (2011). To measure the timescale of an area we scrambled a natural speech at the “words” timescale ( $0.7 \pm 0.5$  s), “sentences” timescale ( $7.7 \pm 3.5$  s), and “paragraphs” timescale ( $38.1 \pm 17.6$  s), and then used a novel neural unscrambling procedure. In the procedure we unscrambled (reordered) the neural responses to each individual sentence (paragraph) that was presented in the scrambled sentence (paragraph) condition and we compared the unscrambled responses to the responses in the intact (100%) condition. In addition, we compared the responses to the intact story with time-reversed responses to the “backward” story (“reverse backward”). The unscrambling procedure revealed that the timescale of processing gradually increases from early sensory areas to higher order perceptual and cognitive areas. Early sensory cortices such as the primary auditory cortex (A1+) have relatively short TRWs (up to hundreds of milliseconds) and therefore respond similarly to each unit of information regardless of the scrambling level. In contrast, high order areas such as the TPJ responded differently to the same set of sentences or paragraphs when they were presented in a different order, as one would expect for an area in which the momentary response is dependent on the stimulus history over long timescales. Areas at intermediate levels of the hierarchy (e.g., pSTG) exhibited a response profile consistent with a temporal integration window of intermediate length (Lerner et al. 2011). \* $P < 0.05$ ; \*\* $P < 0.005$ ; \*\*\* $P < 0.0005$ .

effect may be more apparent for more extreme dilations (e.g., 400%).

What causes the bottleneck in processing for the speeded speech? Notably, speech intelligibility can be recovered even for triple speed speech if short (e.g., 80 ms) periods of silence are interspersed between each short (e.g., 40 ms) unit of speeded speech (Ghitza and Greenberg 2009). The fact that this manipulation recovers speech intelligibility indicates that the earliest processing regions (sensitive to acoustic features <100 ms in duration) are able to process the individual 40-ms speech units. Thus it must be the higher order regions (with intermediate and longer TRWs) that were unable to perform their

integrative function at triple speed and that recovered this ability when the pace was lowered via the insertion of silence.

Limitations on neural response scaling capabilities were also reported by Vagharchakian et al. (2012) who identified higher order language regions that were unable to accelerate their processing in response to accelerated speech. However, in this study we demonstrated that sensory areas as well as higher order areas rescale their responses to match the incoming information rate. This result is inconsistent with the model of Vagharchakian et al. (2012), who argued that higher stages of language processing operate at a fixed speed. There are two possible explanations for the discrepancy. First, the study of

Vagharchakian et al. (2012) used short (2.8 s) isolated sentences in an event-related design. Such a design may be suboptimal for detecting rescaling of the underlying neural responses in higher order areas given that individual sentences do not drive higher order language responses in the same way as a continuous narrative (Xu et al. 2005; Lerner et al. 2011). Second, it is possible that different aspects of neural activity are emphasized by trigger-average analyses and by continuous correlation analyses (see Ben-Yakov et al. 2012).

Our behavioral and neural measurements point to the conclusion that the brain does not code information in absolute temporal coordinates. Rather, linguistic and extra-linguistic circuits can flexibly scale their response dynamics to match the temporal rate of their inputs, just as they scale their spatial receptive fields to match informational density over space. The level of flexibility most likely reflects the range of scales, at which ecologically relevant information arrives from the environment.

#### ACKNOWLEDGMENTS

We thank Janice Chen, Rava A. da Silveira, Lauren Silbert, Alana D'Alfonso, Yaara Y. Yeshurun-Dishon, Nava Rubin, and Miki Fukui for helpful comments on the manuscript.

Present address of Y. Lerner: Tel Aviv Sourasky Medical Center, 6 Weizmann St., Tel-Aviv 64239, Israel.

Present address of M. Katov: Neurobiology Dept., Weizmann Institute of Science, Rehovot 76100, Israel.

#### GRANTS

This work was supported by the National Institute of Mental Health Grant R01-MH-094480 (to U. Hasson and C. J. Honey).

#### DISCLOSURES

**AQ: 3** No conflicts of interest, financial or otherwise, are declared by the author(s).

#### AUTHOR CONTRIBUTIONS

Author contributions: Y.L. and U.H. conception and design of research; Y.L. performed experiments; Y.L., C.J.H., and M.K. analyzed data; Y.L., C.J.H., M.K., and U.H. interpreted results of experiments; Y.L., C.J.H., and M.K. prepared figures; Y.L. drafted manuscript; Y.L., C.J.H., M.K., and U.H. edited and revised manuscript; U.H. approved final version of manuscript.

#### REFERENCES

- Adank P, Devlin JT.** On-line plasticity in spoken sentence comprehension: adapting to time-compressed speech. *Neuroimage* 49: 1124–1132, 2010.
- Ahissar E, Nagarajan S, Ahissar M, Protopapas A, Mahncke H, Merzenich MM.** Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. *Proc Natl Acad Sci USA* 98: 13367–13372, 2001.
- Beasley DS, Bratt GW, Rintelmann WF.** Intelligibility of time-compressed sentential stimuli. *J Speech Hear Res* 23: 722–731, 1980.
- Ben-Yakov A, Honey CJ, Lerner Y, Hasson U.** Loss of reliable temporal structure in event-related averaging of naturalistic stimuli. *Neuroimage* 63: 501–506, 2012.
- Benjamini Y, Hochberg Y.** Controlling the false discovery rate—a practical and powerful approach to multiple testing. *J Roy Stat Soc B Met* 57: 289–300, 1995.
- Benjamini Y, Yekutieli D.** The control of the false discovery rate in multiple testing under dependency. *Ann Stat* 29: 1165–1188, 2001.
- Chodorow MS.** Time-compressed speech and the study of lexical and syntactic processing. In: *Sentence Processing: Psycholinguistics Studies Presented to Merrill Garrett*, edited by Cooper WE, Walker EC. Hillsdale, NJ: Erlbaum, 1979, p. 87–111.
- Foulke E, Sticht TG.** Review of research on the intelligibility and comprehension of accelerated speech. *Psychol Bull* 72: 50–62, 1969.
- Foulke E.** The perception of time compressed speech. In: *The Perception of Language*. Horton DL, Jenkins JJ. Columbus, OH: Merrill, p. 79–107, 1971.
- Friederici AD.** Towards a neural basis of auditory sentence processing. *Trends Cogn Sci* 6: 78–84, 2002.
- Friederici AD.** The cortical language circuit: from auditory perception to sentence comprehension. *Trends Cogn Sci* 16: 262–268, 2012.
- Ghitza O, Greenberg S.** On the possible role of brain rhythms in speech perception: intelligibility of time-compressed speech with periodic and aperiodic insertions of silence. *Phonetica* 66: 113–126, 2009.
- Glover GH.** Deconvolution of impulse response in event-related BOLD fMRI. *Neuroimage* 9: 416–429, 1999.
- Gütig R, Sompolinsky H.** Time-warp-invariant neuronal processing. *PLoS Biol* 7: e1000141, 2009.
- Hasson U, Malach R, Heeger DJ.** Reliability of cortical activity during natural stimulation. *Trends Cogn Sci* 14: 40–48, 2010.
- Hasson U, Nir Y, Levy I, Fuhrmann G, Malach R.** Intersubject synchronization of cortical activity during natural vision. *Science* 303: 1634–1640, 2004.
- Hasson U, Yang E, Vallines I, Heeger DJ, Rubin N.** A hierarchy of temporal receptive windows in human cortex. *J Neurosci* 28: 2539–2550, 2008.
- Hertrich I, Dietrich S, Ackermann H.** How can audiovisual pathways enhance the temporal resolution of time-compressed speech in blind subjects? *Front Psychol* 4: 530, 2013.
- Hickok G, Poeppel D.** Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition* 92: 67–99, 2004.
- Hickok G, Poeppel D.** The cortical organization of speech processing. *Nat Rev Neurosci* 8: 393–402, 2007.
- Honey CJ, Thesen T, Donner TH, Silbert LJ, Carlson CE, Devinsky O, Doyle WK, Rubin N, Heeger DJ, Hasson U.** Slow cortical dynamics and the accumulation of information over long timescales. *Neuron* 76: 423–434, 2012.
- Lerner Y, Honey CJ, Silbert LJ, Hasson U.** Topographic mapping of a hierarchy of temporal receptive windows using a narrated story. *J Neurosci* 31: 2906–2915, 2011.
- Li W, Piech V, Gilbert CD.** Contour saliency in primary visual cortex. *Neuron* 50: 951–962, 2006.
- Luo H, Poeppel D.** Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron* 54: 1001–1010, 2007.
- Miller JL, Grosjean F, Lomanto C.** Articulation rate and its variability in spontaneous speech: a reanalysis and some implications. *Phonetica* 41: 215–225, 1984.
- Mukamel R, Gelbard H, Arieli A, Hasson U, Fried I, Malach R.** Coupling between neuronal firing, field potentials, and fMRI in human auditory cortex. *Science* 309: 951–954, 2005.
- Mukamel R, Nir Y, Harel M, Arieli A, Malach R, Fried I.** Invariance of firing rate and field potential dynamics to stimulus modulation rate in human auditory cortex. *Hum Brain Mapp* 32: 1181–1193, 2011.
- Nourski KV, Reale RA, Oya H, Kawasaki H, Kovach CK, Chen H, Howard MA, Brugge JF.** Temporal envelope of time-compressed speech represented in the human auditory cortex. *J Neurosci* 29: 15564–15574, 2009.
- Obleser J, Eisner F, Kotz SA.** Bilateral speech comprehension reflects differential sensitivity to spectral and temporal features. *J Neurosci* 28: 8116–8123, 2008.
- Op De Beeck H, Vogels R.** Spatial sensitivity of macaque inferior temporal neurons. *J Comp Neurol* 426: 505–518, 2000.
- Peelle JE, McMillan C, Moore P, Grossman M, Wingfield A.** Dissociable patterns of brain activity during comprehension of rapid and syntactically complex speech: evidence from fMRI. *Brain Lang* 91: 315–325, 2004.
- Peelle JE, Gross J, Davis MH.** Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. *Cereb Cortex* 23: 1378–1387, 2013.
- Poldrack RA, Temple E, Protopapas A, Nagarajan S, Tallal P, Merzenich M, Gabrieli JD.** Relations between the neural bases of dynamic auditory processing and phonological processing: evidence from fMRI. *J Cogn Neurosci* 13: 687–697, 2001.
- Privman E, Nir Y, Kramer U, Kipervasser S, Andelman F, Neufeld MY, Mukamel R, Yeshurun Y, Fried I, Malach R.** Enhanced category tuning revealed by intracranial electroencephalograms in high-order human visual areas. *J Neurosci* 27: 6234–6242, 2007.

- Rauschecker JP, Scott SK.** Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nat Neurosci* 12: 718–724, 2009.
- Romanski LM, Averbeck BB.** The primate cortical auditory system and neural representation of conspecific vocalizations. *Annu Rev Neurosci* 32: 315–346, 2009.
- Sheinberg DL, Logothetis NK.** Noticing familiar objects in real world scenes: the role of temporal cortical neurons in natural vision. *J Neurosci* 21: 1340–1350, 2001.
- Smith BL.** Effects of speaking rate on temporal patterns of English. *Phonetica* 59: 232–244, 2002.
- Summerfield Q.** Articulatory rate and perceptual constancy in phonetic perception. *J Exp Psychol* 7: 1074–1095, 1981.
- Talairach J, Tournoux P.** *Co-Planar Stereotaxic Atlas of the Human Brain*. New York: Thieme Medical Publishers, 1988.
- Theiler J, Eubank S, Longtin A, Galdrikian B, Farmer J.** Testing for nonlinearity in time series: the method of surrogate data. *Physica D Nonlinear Phenomena* 58: 77–94, 1992.
- Vagharchakian L, Dehaene-Lambertz G, Pallier C, Dehaene S.** A temporal bottleneck in the language comprehension network. *J Neurosci* 32: 9089–9102, 2012.
- Wingfield A.** Acoustic redundancy and the perception of timecompressed speech. *J Speech Hear Res* 18(1): 96–104, 1975.
- Wingfield A.** Cognitive factors in auditory performance: context, speed of processing and constraints of memory. *J Am Acad Audiol* 7: 175–182, 1996.
- Wingfield A, Peelle JE, Grossman M.** Speech rate and syntactic complexity as multiplicative factors in speech comprehension by young and older adults. *Aging Neuropsychol C* 10: 310–322, 2003.
- Xu J, Kemeny S, Park G, Frattali C, Braun A.** Language in context: emergent features of word, sentence, and narrative comprehension. *Neuroimage* 25: 1002–1015, 2005.
- Zipser K, Lamme VA, Schiller PH.** Contextual modulation in primary visual cortex. *J Neurosci* 16: 7376–7389, 1996.

