Perceiving and remembering speech depend on multifractal nonlinearity in movements producing and exploring speech

Lauren Bloomfield[1], Elizabeth Lane[1], Madhur Mangalam[2] and Damian G. Kelty-Stephen[1]

[1]Department of Psychology, Grinnell College, Grinnell, IA 50112, USA

[2]Department of Physical Therapy, Movement and Rehabilitation Sciences, Northeastern University, Boston, MA 02115, USA

Author for correspondence:

Damian G. Kelty-Stephen

e-mail: keltysda@grinnell.edu

Ethics statement. All participants provided informed consent approved by the Institutional Review Board (IRB) at Grinnell College, Grinnell, IA.

Data accessibility. All data analyzed in the present study are available as supplementary material.

Author contributions: L.B., E.L., and D.G.K-S. conceived and designed research; L.B., E.L., and D.G.K-S. performed experiments; L.B., E.L., M.M., and D.G.K-S. curated and analyzed data; L.B., E.L., M.M., and D.G.K-S. interpreted results of experiments; M.M. and D.G.K-S. prepared figures; L.B., E.L., M.M., and D.G.K-S. drafted manuscript; L.B., E.L., M.M., and D.G.K-S. edited and revised manuscript; L.B., E.L., M.M., and D.G.K-S. approved final version of manuscript.

Competing interests. The authors have no competing interests to declare.

Supplementary materials

Supplementary Note S1. Reading comprehension test given to each participant.

Supplementary Dataset S1. Data modeled in the Poisson regression of self-paced word RT (section 3.1).

Supplementary Dataset S2. Data modeled in the Poisson regression of cumulative correct answers in the post-narrative test (section 3.2).

Abstract

Speech perception and memory for speech require active engagement. Gestural theories have emphasized mainly the effect of speaker's movements on speech perception. They fail to address the effects of listener movement, focusing on communication as a boundary condition constraining movement among interlocutors. The present work attempts to break new ground by using multifractal geometry of physical movement as a common currency for supporting both sides of the speaker-listener dyads. Participants self-paced their listening to a narrative, after which they completed a test of memory querying their narrative comprehension and their ability to recognize words from the story. The multifractal evidence of nonlinear interactions across timescales predicted the fluency of speech perception. Self-pacing movements that enabled listeners to control the presentation of speech sounds constituted a rich exploratory process. The multifractal nonlinearity of this exploration supported several aspects of memory for the perceived spoken language. These findings extend the role of multifractal geometry in the speaker's movements to the narrative case of speech perception. In addition to posing novel basic research questions, these findings make a compelling case for calibrating multifractal structure in text-to-speech synthesizers for better perception and memory of speech.

Keywords: acoustic; auditory perception; complexity; learning disability; multifractality; nonlinear

1. Introduction

Listening to speech entails active engagement. Even when listening is most passive (e.g., when listening to a podcast on headphones), beyond vibrations from the eardrums, the listener consults the lexicon, a trove of possible words and their corresponding meaning [1]. The listener must explore the characteristics of the sounds themselves. Listeners can potentially face an uncertain terrain when exploring the speech around them. The present study investigates how we may use movement to explore speech that supports perception of and cognition about spoken words.

Speech perception depends on the articulatory motor processes shaping speech's acoustic features [2–4]. Motor processes reshape phonemes in context- and sequence-dependent ways (e.g., in coarticulation) that text-to-speech synthesis has struggled to emulate [5]. Speech sounds reflect movements beyond articulators to distal parts of the body [6]—potentially explaining their utility for diagnosing Parkinson's disease [7]. However, movement is now being recognized as an aspect of speech that otherwise passive listeners can use [8]. Gestural theories have been at most ambivalent about the role of listener's actions in perception and cognition of spoken words (e.g., compare [9] with [10,11]), most recently focusing on communication as a boundary condition constraining movement among interlocutors [12]. However, perceiving speech becomes more difficult when we listen to movement systems that act less like we do ourselves, as when we listen to a speaker from another linguistic group [13].

Here, we test an experimental contrast of a human speaker with the text-to-speech synthesis in a between-groups design. Text-to-speech synthesizers may lack many context- and sequence-dependencies in human speech [14], and they dramatically reshape the movement-structured terrain for the listener. For present purposes, we use the speech waveform as an operationalization of speech movements. Although it is not an anatomically specific measurement of individual speech articulators, the speech waveform carries the imprint of fine kinetic and kinematic markers of speech articulators [15], subtle enough to aid in distinguishing phonation type [16], misarticulated stops from cleft lip or palate [17], and dysarthrias [18].

The geometry of listener movements for exploring speech sounds may be no less crucial than speaker movements. Auditory perception has strong roots in a listener's bodywide capability for physical movement [19–25], and spoken-language perception depends on movement, e.g., with activity of the motor cortex in covert imitation of speech sounds supporting interactive alignment across many scales of dyadic conversation [26]. For instance, even in the case of listening to recorded speech, the spatial organization of speech in the space around the listener's body facilitates spoken-word perception under difficult listening circumstances, for example, with noise [27] or reverberation [28]. We can elicit overt exploratory behaviors, such as key pressing, during speech perception through the "self-paced listening" paradigm requiring participants to make key presses to signal readiness for consecutive words in a sequence [29]. Self-pacing affords the reader more control in exploring linguistic stimuli, with progress or pause indicating the waxing and waning of focus and processing, and giving readers with attentional lapses or sensory deficits a greater opportunity to allocate attention and memory for speech stimuli [30,31].

We align articulatory movement for speech production reflected by the speech waveform with exploratory self-pacing movements for speech perception within a single geometric formalism: multifractal geometry. Multifractal geometry may reveal a causal relationship between (1) motor processes underlying speech stimuli and (2) listener movements for exploring those stimuli. The multifractal structure might predict how well individual listeners process the spoken-word stimulus (figure 1*a*, top right) and later exhibit their memory for the spoken words (figure 1*a*, bottom left). In Part-1, participants listened to a story, word-by-word, articulated by either human speech or text-to-speech synthesizer. This manipulation allowed variations in speech-stream geometry across different words and movement systems (i.e., natural human speech vs. computer-simulated analog). In Part-2, participants completed a memory test querying their comprehension of and memory for words from the story.

1.1. Multifractal geometry helps to formalize the movement-based support for conveying information

Multifractal geometry is a generalization of fractal geometry that encompasses 'multiple' degrees of fractal structure [32–34]. Rare in psycholinguistic research (e.g., [35–37]), variation in fractal structure has begun to

demonstrate usefulness in modeling gestural supports for perceiving speech [3,4] as well as modeling how readers explore language in a narrative [38].

1.2. Multifractality offers statistical estimation of nonlinear interactions across scales, indicating how systems blend current, fleeting behaviors into longer-term behaviors

Multifractal analysis examines how variation depends on time in a measured behavior and how uniformly or unevenly variation unfolds from a brief glimpse to a longer span. Multifractal geometry evaluates variability through a variety of scales within a single measurement. For instance, the analysis segments a time-series measurement, such as audio waveform, into sequential but nonoverlapping bins of the same small size. Each bin represents a short subset of the whole movement at one short time scale. The analysis then estimates variability across those different bins of the same size. Then, multifractal geometry iterates this process for progressively longer timescales.

The average amount of variability in measurement grows as a power law of timescale. Only one power law relates average proportion to timescale for homogeneous time series in which brief and long-term events do not interact. However, for heterogeneous measurements of systems with interactions across scales, a single measurement can contain multiple proportion-to-time power laws. The word 'multifractal' reflects that heterogeneous systems can exhibit multiple power-law relationships each with different fractional exponents. This variety of fractional power-law exponents constitutes a 'multifractal spectrum.'

Two multifractal statistics indicate the strength of relationships propagating brief events across longer timescales. The first metric, the multifractal spectrum width $\Delta\alpha$, is conceptually analogous to standard deviation but reflects a mixture of linear variability and nonlinear variability. The linear variability concerns histogram width and average drift over time. But the nonlinear variability crucially concerns interactions across scales. Another metric 'multifractal nonlinearity' t_{MF} is a t-statistic indicating difference of $\Delta\alpha$ from a sample of expected $\Delta\alpha$ values from strictly linear models of the measurement [39]. Bottom panels of figures 2*b* and 2*c* schematize this difference between linear structure and nonlinear interactions across timescales.

Multifractality is the natural, logical outcome when different timescales interact. Models of speech production/perception include hierarchical models that nest fine, relatively "analog" details of movement and stimulation into progressively broader and more abstract neuroscientific and linguistic categories, structures, or expectations. These models often manifest in Bayesian linear models [40] or convolutional neural networks [41]. These elegant models fall short of explaining the speech-production/perception outcomes in evidence [42]. Their hierarchical structure is so static that as to force symmetrical repetition and mimicry sooner than generating dynamic and time-evolving synergistic structures [43]. The use of recurrence quantification has been a significant step towards examining language use across multiple scales without strictly linear assumptions (e.g., [43]). However, multifractal modeling can address more specifically the role of nonlinearity in cross-scale interactions in producing the asymmetric waxing or waning of specific synergistic coordinations. So, multifractality may answer current calls for formalisms capable of developing falsifiable hypotheses about synergies in language use [13].

Both speech production and language perception exemplify just such nonlinear interactions across timescales. Longer-term speech sequences reshape the brief articulation of a single syllable [5]. Communicative sounds exhibit a range of fractional power-law exponents [44]. Multifractality nonlinearity t_{MF} may predict expressiveness [45]. Speech perception adjusts to compensate for coarticulatory interaction across time [2] and tailors these compensations over longer timescales [46]. Indeed, research has already modeled the effects of the multifractal structure of speech waveforms in supporting speech perception [3,4]. We choose to use the waveform rather than an auditory envelope. Envelopes involve smoothing procedures that may serve multifractal analysis for very long series (e.g., whole bird song phrases [45]). We chose not to use an envelope of any sort because we risked removing finer vibrations in these briefer single-word recordings that would support multifractal modeling over a wider range of scales.

1.3. Multifractal fluctuations in measurements of movement predict how movements support the pick-up, coordination, and later use of information for perceptual judgments

Multifractal geometry has repeatedly captured predictive information about how movement systems loosen their constraints to absorb and then propagate perceptual information. For example, this predictive capacity can be seen in haptic [47,48] and visual perception [49,50], as well as how movement systems integrate information across the body [51,52] and across different modalities [53]. In psycholinguistic domains, greater multifractal nonlinearity in self-paced reading may keep readers' expectations loose and open-minded enough to navigate unexpected plot twists more fluently, (i.e., with a lower increase in subsequent reading-time), indicating a lighter processing load [38].

Multifractal fluctuations support the sharing of information necessary for dexterous behavior across organisms [54,55], just like they do within organisms. Indeed, attention to a speaker engenders a fractal-like 'complexity matching' supporting the listener's comprehension of linguistic content in the signal [56–58]. There has been growing curiosity in how these findings relate to insights from the multifractal formalism [58,59]. We now test whether such multifractal geometry supports this social-information exchange in speech production and perception.

1.4. Two hypotheses: Multifractality nonlinearity supports the uptake and memory for spoken-language

1.4.1. Hypothesis-1: Greater multifractal nonlinearity of speech sounds predicts more fluent response to individual words, prolonging fluency of perceiving text-to-speech synthesized words comparable to human-voice words

We used a self-paced listening task to experimentally test the effect of variations in multifractality—both from word to word within the same speaker and between two speakers with different styles of producing speech. We enlisted a human speaker and a text-to-speech synthesizer to pronounce each word of a narrative separately but in sequence. Despite supporting students with difficulty reading by sight [60], the speech-production in text-to-speech is unfamiliar and perceptually challenging [61]. Because of the potential role of multifractal nonlinearity in supporting expressivity, we predicted that multifractal nonlinearity in speech would speed the word-by-word response in self-paced listening (Hypothesis-1a). Likewise, we predicted that multifractal nonlinearity would support the fluency of text-to-speech synthesis perception and defer any gradual slowing of the self-pacing in response to text-to-speech speech relative to human speech (Hypothesis-1b). We explored effects of raw multifractal spectrum width without anticipating any single direction.

1.4.2. Hypothesis-2: Greater multifractal spectrum width and greater nonlinearity of the self-paced key press interval time series would predict stronger performance on post-narrative test of comprehension and recognition memory

We predicted that greater multifractal nonlinearity across key-press sequences would predict better performance on a post-narrative memory test. We explored effects of raw multifractal spectrum width without anticipating any single direction.

2. Materials and methods

2.1. Participants

Nine healthy men and eleven healthy women (*mean±s.d.* age, 20.10±1.29 years) participated after providing informed consent. Three participants reported having a learning disability, two with attention deficit hyperactivity disorder (ADHD) and 1 with dyslexia, amounting to the same proportion of people with the same diagnoses in the United States population [62,63].

2.2. Experimental task and procedure

Participants were randomly assigned to hear the voice of either an adult woman or Acapela's U.S. English text-to-speech female voice 'Sharon' (Acapela Inc., Mons, Belgium) using iPad app 'Voice Dream.' Both produced speech recordings of 2,027 words in sequence from *The Atlantic* article "Torching the Modern-Day Library of Alexandria." The text-to-speech recording had two fewer words because of a text-parsing error (e.g., omitting to

convert "$" symbols to the word "dollars" after the amount). Human speech and text-to-speech both produced words interspersed with pauses to allow parsing. Words were not controlled to have the same pitch and duration between speakers, as the primary feature of interest was the multifractal nonlinearity of the speaker's word production. Hence, the within-speaker assessment of nonlinearity was the higher priority and controlled for average pitch and duration effects through comparison to surrogates built to mimic the amplitude spectrum and duration. Multiple standard psycholinguistic features of these words were assessed: word frequency, 'logWF,' in SUBTLEX [64]; and number of orthographic and phonological neighbors—'OrthN' and 'PhonN,' respectively—in CLEARPOND [65].

E-Prime software (Psychology Software Tools Inc., Pittsburgh, PA) presented individual audio recordings of each word in original sequence through headphones. Participants sat at an E-Prime-ready computer and were instructed: "Listen to the audio stimuli and press the spacebar after you feel as though you have understood the word you just heard. Try to pay attention to the passage because comprehension and word-memory questions will be asked at the end of the experiment. However, if you miss a word, do not worry and continue to move on because you cannot go back." Word recordings played only once immediately as the spacebar was released. Participants could press the spacebar again, either while or after the recording played, to hear the next word, potentially interrupting the full word before its completion in order to hear the next word. E-Prime recorded the response time (RT) in milliseconds (up to one decimal place) from each word onset to subsequent button press. After the completion of the self-paced listening task, participants completed a 20-item pen-and-paper test (see Reading Comprehension Test in Supplementary Note S1). The first ten items prompted participants to indicate whether each of the ten sentences was 'true' or 'false' based on their comprehension of the narrative. The later ten items prompted participants to demonstrate word-recognition memory by indicating whether each of 10 words appeared in the text. Only five of the words appeared in the text; we made seven of the sentences true so as not to give participants the chance to guess that both tests should have the same number of incorrect items.

2.3. Multifractal metrics

We computed multifractal metrics for the absolute value of each spoken word's audio waveform (whether from human speech or text-to-speech) and each participant's RT sequence. The following summary of the metrics is brief and mathematically abbreviated. We invite readers interested in a longer, more descriptive explanation to examine an open-access resource here [66] for both a more conversational introduction and R syntax for estimating these metrics.

2.3.1. Multifractal spectra width

Chhabra and Jensen's direct method [67] samples non-negative series $u(t)$ at progressively larger scales such that proportion of signal $P_i(L)$ falling within the ith bin of scale L is

$$P_i(L) = \frac{\sum_{k=(i-1)L+1}^{iL} u(k)}{\sum u(t)} \quad (1)$$

As L increases, $P_i(L)$ represents progressively larger proportion of $u(t)$,

$$P(L) \propto L^{\alpha} \quad (2)$$

suggesting growth of proportion according to 'singularity' strength α. $P(L)$ exhibits multifractal dynamics when it grows heterogeneously across timescales L according to multiple potential fractional singularity strengths, such that

$$P_i(L) \propto L^{\alpha_i} \quad (3)$$

whereby each ith bin may show a distinct relationship of $P(L)$ with L. The spectrum of singularities is itself the multifractal spectrum, and its width $\Delta\alpha(\alpha_{max} - \alpha_{min})$ indicates the heterogeneity of these relationships between proportion and timescale [32,68].

This method estimates $P(L)$ for N_L nonoverlapping bins of L-sizes and accentuates higher or lower $P(L)$ by using parameter $q > 1$ and $q < 1$, respectively, as follows

$$\mu_i(q, L) = \frac{[P_i(L)]^q}{\sum_{i=1}^{N_L}[P_i(L)]^q}. \quad (4)$$

See *figure* 2.

$\alpha(q)$ is the singularity for $\mu(q)$-weighted $P(L)$ estimated by

$$\alpha(q) = -\lim_{L\to\infty}\frac{1}{lnL}\sum_{i=1}^{N}\mu_i\,(q, L)lnP_i(L)$$

$$\lim_{L\to 0}\frac{1}{lnL}\sum_{i=1}^{N}\mu_i\,(q, L)lnP_i(L). \quad (5)$$

Estimates $\alpha(q)$ belong to the multifractal spectrum if Shannon entropy of $\mu(q, l)$ scales with L according to a dimension $f(q)$, where

$$f(q) = -\lim_{L\to\infty}\frac{1}{lnL}\sum_{i=1}^{N}\mu_i\,(q, L)ln\mu_i(q, L)$$

$$\lim_{L\to 0}\frac{1}{lnL}\sum_{i=1}^{N}\mu_i\,(q, L)ln\mu_i(q, L). \quad (6)$$

For q generating scaling relationships (Eqs. 5 and 6) with correlation coefficient, $r > 0.95$, the parametric curve $\big(\alpha(q), f(q)\big)$ or $\big(\alpha, f(\alpha)\big)$ constitutes the multifractal spectrum with width $\Delta\alpha = \alpha_{max} - \alpha_{min}$ (figure 2, bottom right).

2.3.2. Multifractal nonlinearity

To estimate how much observed multifractality reflects nonlinearity [69], $\Delta\alpha$ of each original series was compared to $\Delta\alpha$ of 32 surrogate series obtained using Iterated Amplitude Adjusted Fourier Transformation (IAAFT) [39]. IAAFT generates surrogates that randomize phase ordering of the series' spectral amplitudes while preserving only linear temporal correlations. We defined multifractal nonlinearity as the one-sample *t*-statistic (henceforth, t_{MF}) comparing $\Delta\alpha$ of the original series to that of the surrogates (figure 3).

2.4. Statistical analysis

To test Hypothesis-1, a Poisson regression of self-paced word RT using glmer() in the package *lme4* [70] for *R* [71] included effects of logarithmic sound length (lnSL), word frequency (logWF), number of phonological neighbors (PhonN), number of orthographic neighbors (OrthoN), text-to-speech (TTS = 1 and 0 for synthesizer and human speech, respectively), and t_{MF}. It also included two time predictors spanning trials across the experiment: word# (word number incrementing by 1 with each word) and, because RT shows a general power-law decrease across progressive words for self-paced processing of narrative text [72], 1/word#. The model included the interaction effects of TTS×word#, as well the interactions 1/word#×TTS×logWF, 1/word#×TTS×PhonN, 1/word#×TTS×OrthN, 1/word#×TTS×t_{MF}, in addition to all other lower-order interactions. Additional terms or interactions either were nonsignificant or led to poor model convergence (e.g., $\Delta\alpha$). We used Poisson regression because E-Prime records RT in discrete units (i.e., tenths of a millisecond). Poisson regression is regularly used to model similarly discrete time-to-events [73–77].

To test Hypothesis-2, a Poisson regression of cumulative correct answers in the post-narrative test included effects of test type (WRM = word recognition memory as opposed to comprehension), condition (TTS = having heard text-to-speech versus human speech), orthogonal cubic polynomials of questions 1 through 10 (calculated using the *R*-function poly() [71]) and various metrics on individual participants' self-paced RT series including

linear features (mean and median) and multifractal metrics ($\Delta\alpha$ and t_{MF}). Modeling tested typeWRM×conditionTTS×RT-$\Delta\alpha$, typeWRM×conditionTTS×RT-t_{MF}, conditionTTS×log(RT-mean), conditionTTS×log(RT-median), as well as question×typeWRM and question×$\Delta\alpha$, and all lower-order interactions. This model also included a dichotomous predictor 'LD-status' (1 or 0 = yes or no) to control any benefits that a learning disability might experience due to the self-paced format [30,31] or their greater fluency with text-to-speech synthesis [78]. Note that LD-status failed to improve the model for Hypothesis 1 and so was omitted from that model. The pattern of significant effects were the same with or without LD-status.

Both models used 1,000 resamplings to generate bootstrapped standard errors, *p*-values, and power estimates. All significant predictors showed at least 80% statistical power (i.e., > 800 resamplings yielded nonzero coefficients in the same direction as the original data) [79].

3. Results

Figure 4 shows illustrative output from multifractal analysis of both speech and RT sequences.

3.1. Model of word-by-word RT

RT for human speech (*mean* = 599.46, *s.e.m.* = 2.17) was lower than for text-to-speech (*mean* = 654.80, *s.e.m.* = 2.44). Table 1 lists this model's coefficients.

3.1.1. Testing Hypothesis-1a: Multifractal nonlinearity sped the response in self-paced listening

Speech stimuli with greater multifractal nonlinearity evoked faster responses (t_{MF}: $B = -2.77\times10^{-5}$, $p < 0.01$), although the difference was small, ranging from 0.90% to 2.74% as multifractal nonlinearity varied from 1st to 3rd quartiles (table 1). We found comparable effects of word frequency (logWF: $B = -3.22\times10^{-2}$, $p < 0.0001$) and phonological neighborhood size (PhonN: $B = -3.97\times10^{-4}$, $p < 0.05$) in speech perception. RT increased with logarithmically-scaled sound length (lnSL: $B = 3.44\times10^{-1}$, $p < 0.0001$): it takes longer to listen to longer sounds.

3.1.2. Testing Hypothesis-1b: Multifractal nonlinearity deferred the slower response to text-to-speech synthesis—later than other psycholinguistic features did

Word-by-word RTs decreased at a power law (e.g., [72]; 1/word#: $B = 1.02\times10^{0}$, $p < 0.001$). Word-by-word RT increased linearly with each TTS word (TTS: $B = -7.16\times10^{-2}$, $p < 0.0001$; TTS×word#: $B = 1.42\times10^{-4}$, $p < 0.0001$; figure 5*a*). However, multifractal nonlinearity of speech sounds predicted a stronger power-law decay in RT in the TTS case (1/word#×TTS×t_{MF}, $B = 1.81\times10^{-3}$, $p < 0.01$). Consequently, response to text-to-speech became significantly slower only much later for sounds with greater multifractal nonlinearity, with significant differences between RT for TTS and RT for human speech appearing on word #1140 for low t_{MF} (1st quartile; thin gray vertical line figure 5*a*) and on word #1280 for high t_{MF} (3rd quartile; thick gray vertical line in figure 5*a*). Greater multifractal nonlinearity led to comparable RTs for human speech and text-to-speech speech comparable for 140 words, or 12.28% longer than lower multifractality.

Effects of multifractal nonlinearity held above and beyond effects of standard psycholinguistic features. Phonological neighborhood yielded an even stronger power-law decay with text-to-speech (1/word#×TTS×PhonN: $B = 8.21\times10^{-2}$, $p < 0.001$); figure 5*b*), prolonging comparable RTs for text-to-speech and for human speech until about half way through the text. Frequency showed a comparable effect (TTS×logWF: $B = 1.65\times10^{-2}$, $p < 0.0001$; 1/word#×logWF: $B = 2.84\times10^{-1}$, $p < 0.0001$; figure 5*c*). Finally, orthographic neighborhood size exhibited no significant effect ($p > 0.05$), likely reflecting the auditory rather than the visual form of the task.

3.2. Testing Hypothesis 2: Model of cumulative correct responses on the 20-item two-part memory test

Story comprehension and word recognition for human speech (*mean* = 7.3 and 5.1 out of 10, *s.e.m.* = 0.45 and 0.46, respectively) were comparable to the same scores for text-to-speech (*mean* = 7.2 and 4.8 out of 10, *s.e.m.* = 0.39 and 0.66, respectively). Table 2 lists this model's coefficients. Word recognition was easier ($B = 2.99\times10^{-}$

^{1}, $p < 0.05$), and text-to-speech synthesizers might have supported more effective learning with self-pacing ($B = 4.25\times10^{0}$, $p < 0.01$), whereas controlling the pacing of human speech might have been distracting.

Multifractality of RT from linear sources (i.e., $\Delta\alpha$) diminished memory performance only in word-recognition memory in the human-speech case ($B = -4.66\times10^{0}$, $p < 0.0001$; figure 6*b*). However, the text-to-speech manipulation showed smaller negative effects of multifractality from linear sources on both word-recognition memory and on narrative-comprehension memory (typeWRM×conditionTTS×RT-Δα: $B = 2.53\times10^{0}$, $p < 0.05$; conditionTTS×RT-Δα: $B = -2.28\times10^{0}$, $p < 0.05$; table 2). Multifractal nonlinearity reduced narrative-comprehension memory ($B = -7.11\times10^{-3}$, $p < 0.05$; figure 7*a*) but promoted word-recognition memory following self-paced listening to human speech ($B = 2.81\times10^{-2}$, $p < 0.0001$; figure 7*b*). From 1st to 3rd quartiles of RT-t_{MF}, narrative-comprehension memory became 5% to 22% less accurate, and word-recognition memory became 25% to 150% more accurate (table 2). Memory performance on both tests decreased for participants with greater multifractal nonlinearity in text-to-speech condition only in word recognition ($B = -3.13\times10^{-2}$, $p < 0.0001$), but linear sources of multifractality promoted word-recognition memory.

Over the course of each 10-question memory test, each correct response had the potential to prime more memory for the text. Correct responses increased at a linear relationship with question number (poly(Question, 3)1: $B = 8.80\times10^{0}$, $p < 0.0001$; figure 6). Multifractality from linear sources accentuated only the linear growth of correct responses with question number (i.e., $\Delta\alpha$; poly(Question,3)1×RT-$\Delta\alpha$: $B = 16.65\times10^{0}$, $p < 0.01$; figure 6). Multifractal nonlinearity did not show significant effects (table 2; figure 7). Correct responses in the word-recognition test increase across the 10 questions with a stronger cubic relation, including a negative quadratic relationship (poly(Question, 3)2×typeWRM: $B = -3.17\times10^{0}$, $p < 0.0001$; poly(Question, 3)3×typeWRM: $B = 1.16\times10^{0}$, $p < 0.05$; figures 6 and 7).

Post-text memory for the text showed a minor but significant decrease with longer RTs during the text for human speech (log(RT-mean): $B = 2.46\times10^{0}$, $p < 0.0001$; log(RT-median): $B = -2.48\times10^{0}$, $p < 0.0001$) and for text-to-speech (TTS×log(RT-mean): $B = -4.12\times10^{0}$, $p < 0.0001$; TTS×log(RT-median): $B = 3.57\times10^{0}$, $p < 0.0001$). Considering logarithm of mean and median RT together, longer (3rd quartile) RTs predicted 4% more accurate memory performance than shorter RTs (1st quartile) in human speech. The effect was smaller and reverse in the text-to-speech condition: 0.3% less accurate memory performance (table 2).

3.2.1 Clarifying the Hypothesis 2 result: Controlling for LD-status effect indicates that multifractal nonlinearity only reduces narrative-comprehension in text-to-speech condition

We ran a second version of this model that included LD-status to control for the fact that readers with learning disabilities can benefit from self-pacing [31] and from their fluency using text-to-speech synthesis [78]. This predictor both improved model fit ($\chi^2(1) = 4.65$, $p < 0.05$) and demonstrated greater than 80% power in a 1000 bootstrapped resamplings of the data. This reanalysis confirmed that participants with learning disability did in fact remember more from the text ($B = 3.75\times10^{-1}$, $p < 0.0001$). Furthermore, including the LD-status predictor led to different estimates for main effects of RT multifractality and its interaction with condition. Specifically, including the LD-status predictor clarified that multifractal nonlinearity had no main effect, that the negative main effect actually belonged to multifractal linearity (i.e., RT-$\Delta\alpha$: $B = -2.31\times10^{0}$, $p < 0.01$), and that multifractal nonlinearity only predicted reduced memory in the text-to-speech condition (TTS×RT-t_{MF}: $B = -1.37\times10^{-2}$, $p < 0.05$).

4. Discussion

We tested two hypotheses about the effects of multifractal nonlinearity t_{MF} in perception and memory for spoken language. Hypothesis-1 was that greater t_{MF} in speech stimuli would make speech perception more fluent, particularly in keeping responses text-to-speech synthesis as fluent as responses to human speech. Hypothesis-2 was that greater t_{MF} in the RT sequence in self-paced listening to a narrative would support better memory of the speech stimuli. Results supported Hypothesis-1 in full but Hypothesis-2 only in the word-recognition part of the memory test and not in the narrative-comprehension portion of the memory test. These findings carry implications not only for basic research into the movement basis for speech perception and memory for speech but also for the development of more effective text-to-speech synthesis for students with learning disabilities.

4.1. Multifractal nonlinearity promotes quicker response to speech stimuli and processing of text-to-speech synthesis later than traditional psycholinguistic features do

Multifractal nonlinearity of speech was associated with more fluent speech perception across a narrative. Whether the speaker was human or machine, the inevitable interactions across timescales as movement systems produce speech sounds (human speech and text-to-speech) promoted spoken-word perception in narrative sequence. Not only do articulatory motor processes matter to perception [8–11,80,81], their perceptual relevance goes deeper than linguistic distinctions (e.g., fricatives or stops and voice-onset times) and is rooted in more generic nonlinear dynamics with no currently recognized linguistic meaning [3,4]. The absence of any effects of multifractal spectrum width suggests that this aspect of speech perception depends on nonlinear and not linear relationships across timescales.

Multifractality exerted its role in supporting fluent processing of synthesized speech later than frequency and phonological neighborhoods, but longer than frequency supported. This result likely reflects that multifractal nonlinearity is a deeper attribute of acoustic waveform. Frequency and phonological neighborhood may support the fluent perception of less familiar speech early on because these are high-level features anchored in lexicality (i.e., word status). Meanwhile, multifractal nonlinearity is not specific to words and inheres in the heterogeneous movements producing speech sounds. Perception resorts most clearly to this multifractal movement foundation when linguistic support for word recognition falters.

4.2. Multifractal nonlinearity in self-paced listening to human speech promoted word-recognition memory, and multifractality from linear sources predicted diminished memory on average but increased memory performance as testing continued

Participant's ability to remember the speech stimuli exhibited a more complex relationship between movement systems across the speaker-listener dyad. The two movement systems generating speech (i.e., natural human speech vs. computer-simulated analogue) cultivated distinct aspects of the multifractal structure of self-pacing responses: comprehension of more familiar human speech benefited from nonlinear aspects of multifractal self-pacing behaviors (i.e., t_{MF}), and comprehension of less familiar text-to-speech benefited from linear aspects (i.e., $\Delta\alpha$). Self-paced listening to speech presented word by word may have emphasized individual words at the expense of attention to the narrative's discourse-level aspects. Also, it is possible that including seven true and three false items in the comprehension test reduced variability in responses (e.g., by cuing more memory of the text).

Multifractal nonlinearity appeared at first to diminish narrative-comprehension memory, but controlling for learning-disability advantages with self-paced listening showed instead that multifractal nonlinearity only diminished narrative-comprehension memory in the text-to-speech condition. This point is intriguing because multifractality nonlinearity refers to the interaction of short-range (e.g., word-by-word) scales of activity with long-range (e.g., discourse/narrative) scales of activity. So, it could be that self-paced listening behaviors with this cross-scale interactivity makes maladaptive alignment with text-to-speech synthesis that only produces speech at word-by-word scales.

5. Conclusions

The present work is the latest step in the new enterprise of modeling multifractal effects supporting spoken language use. It addresses current calls in the literature for a straightforward multifractal extension of the notion that a speaker-listener dyad depends on 'complexity matching' between the respective movement systems [56–59]. First, the multifractal evidence of nonlinear interactions across timescales predicts the fluency of speech perception. Second, self-pacing movements that enabled listeners to control the presentation of speech sounds constituted a rich exploratory process. Multifractal nonlinearity of this exploration supports some but not yet all aspects of memory for perceived spoken language. Nonetheless, the present findings suggest an effect of multifractal structure in the both speech production and listener's exploratory movements in speech perception.

Self-paced listening is not ecologically valid for speech perception "in the wild" (e.g., a listener listening to a speaker producing words at the speaker's own preferred pace). However, future work could examine whether

the multifractal structure of postural or head sway could moderate speech perception as it has already been shown to moderate visual and haptic perception (e.g., [47,48,51,52,82]). Even admitting the contrived status of self-paced listening, the present results validate our idea that speech perception and memory reflects an alignment (or not) between the structures of the speaker's and the listener's movements. Speech memory reflected a compromise between the speaker's movement system and the listener's actions' multifractality. The self-pacing paradigm prompting the listener to act to get the next word coaxes some movement whose geometry is also multifractal. Different movement systems can cultivate specifically different (i.e., linear or nonlinear) aspects of the multifractality in self-paced timing. A more familiar speaker (i.e., human speech) sooner leads to multifractal nonlinearity to support speech memory. In contrast, the less familiar speaker (i.e., text-to-speech [14]) sooner leads to linear sources of multifractality to support speech memory.

Future work could investigate aspects of human or text-to-speech voices influencing relative linear or nonlinear effects of multifractality on perception and subsequent memory for speech. For instance, if multifractality influences familiarity of speech, then it could hence enhance text-to-speech synthesis as a learning tool. The present evidence that multifractality of speech movements is essential for speech perception is silent on whether multifractality of speech movements have consequences for memory. Unambiguous comparison of multifractality across dyads was precluded by the radically different data formats between speaker and listener recordings (multiple 44.1 kHz digital recordings vs. single sequences of 2,026 inter-key press intervals). Whether this alignment rests firmly on multifractality could be tested in a 'shadowing' paradigm (i.e., listeners repeating immediately each spoken word, e.g., [83]). It could be beneficial to test how greater familiarity with a particular voice influences the appearance and role of multifractality. Another important point for future work is that words appeared in the narrative context, so it may be worth extending this work to a fuller accounting of RTs not just for words but also for phrases, sentences, paragraphs, etc. [71].

References

1. Aitchison J. 2012 *Words in the Mind: An Introduction to the Mental Lexicon*. Oxford, UK: Wiley-Blackwell.

2. Viswanathan N, Magnuson JS, Fowler CA. 2014 Information for coarticulation: Static signal properties or formant dynamics? *J. Exp. Psychol. Hum. Percept. Perform.* **40**, 1228–1236. (doi:10.1037/a0036214)

3. Hasselman F. 2015 Classifying acoustic signals into phoneme categories: Average and dyslexic readers make use of complex dynamical patterns and multifractal scaling properties of the speech signal. *PeerJ* **3**, e837. (doi:10.7717/peerj.837)

4. Ward RM, Kelty-Stephen DG. 2018 Bringing the nonlinearity of the movement system to gestural theories of language use: Multifractal structure of spoken english supports the compensation for coarticulation in human Speech Perception. *Front. Physiol.* **9**, 1152. (doi:10.3389/fphys.2018.01152)

5. Hill DR, Taube-Schock CR, Manzara L. 2017 Low-level articulatory synthesis: A working text-to-speech solution and a linguistic tool. *Can. J. Linguist.* **62**, 371–410. (doi:10.1017/cnj.2017.15)

6. Pouw W, Paxton A, Harrison SJ, Dixon JA. 2020 Acoustic information about upper limb movement in voicing. *Proc. Natl. Acad. Sci.* **117**, 11364–11367. (doi:10.1073/pnas.2004163117)

7. Tsanas A, Little MA, McSharry PE, Ramig LO. 2011 Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average Parkinson's disease symptom severity. *J. R. Soc. Interface* **8**, 842–855. (doi:10.1098/rsif.2010.0456)

8. Galantucci B, Fowler CA, Turvey MT. 2006 The motor theory of speech perception reviewed. *Psychon. Bull. Rev.* **13**, 361–377. (doi:10.3758/BF03193857)

9. Fowler CA. 1996 Listeners do hear sounds, not tongues. *J. Acoust. Soc. Am.* **99**, 1730–1741. (doi:10.1121/1.415237)

10. Fadiga L, Craighero L, Buccino G, Rizzolatti G. 2002 Speech listening specifically modulates the excitability of tongue muscles: A TMS study. *Eur. J. Neurosci.* **15**, 399–402. (doi:10.1046/j.0953-816x.2001.01874.x)

11. Fowler CA. 2008 The FLMP STMPed. *Psychon. Bull. Rev.* **15**, 458–462. (doi:10.3758/PBR.15.2.458)

12. Fowler CA, Richardson MJ, Marsh KL, Shockley KD. 2008 Language use, coordination, and the emergence of cooperative action. In *Coordination: Neural, Behavioral and Social Dynamics* (eds A Fuchs, VK Jirsa), pp. 261–279. Berlin, Heidelberg: Springer. (doi:10.1007/978-3-540-74479-5_13)

13. Olmstead AJ, Viswanathan N, Cowan T, Yang K. 2021 Phonetic adaptation in interlocutors with mismatched language backgrounds: A case for a phonetic synergy account. *J. Phon.* **87**, 101054. (doi:10.1016/j.wocn.2021.101054)

14. Paul D, Shifas MP V, Pantazis Y, Stylianou Y. 2020 Enhancing speech intelligibility in text-to-speech synthesis using speaking style conversion. *arXiv Prepr. arXiv2008.05809*

15. Gwilliams L, King J-R, Marantz A, Poeppel D. 2020 Neural dynamics of phoneme sequencing in real speech jointly encode order and invariant content. *bioRxiv* , 2020.04.04.025684. (doi:10.1101/2020.04.04.025684)

16. Kadiri SR, Alku P. 2021 Glottal features for classification of phonation type from speech and neck surface accelerometer signals. *Comput. Speech Lang.* **70**, 101232. (doi:10.1016/j.csl.2021.101232)

17. Sudro PN, Vikram CM, Prasanna SRM. 2021 Event-based transformation of misarticulated stops in cleft lip and palate speech. *Circuits, Syst. Signal Process.* (doi:10.1007/s00034-021-01663-3)

18. Narendra NP, Alku P. 2019 Dysarthric speech classification from coded telephone speech using glottal features. *Speech Commun.* **110**, 47–55. (doi:10.1016/j.specom.2019.04.003)

19. Repp BH, Knoblich G. 2007 Action can affect auditory perception. *Psychol. Sci.* **18**, 6–7. (doi:10.1111/j.1467-9280.2007.01839.x)

20. Stoffregen TA, Pittenger JB. 1995 Human echolocation as a basic form of perception and action. *Ecol. Psychol.* **7**, 181–216. (doi:10.1207/s15326969eco0703_2)

21. Wallmeier L, Wiegrebe L. 2020 Self-motion facilitates echo-acoustic orientation in humans. *R. Soc. Open Sci.* **1**, 140185. (doi:10.1098/rsos.140185)

22. Pawlukowska W, Siwczak K, Szylińska A, Turoń A, Rotter I. 2017 The impact of muscle tone on the development of auditory perception – research orientation. *Fam. Med. Prim. Care Rev.* **19**, 382–386. (doi:10.5114/fmpcr.2017.70812)

23. Myers JC, Mock JR, Golob EJ. 2020 Sensorimotor integration can enhance auditory perception. *Sci. Rep.* **10**, 1496. (doi:10.1038/s41598-020-58447-z)

24. Ingber DE. 2006 Cellular mechanotransduction: Putting all the pieces together again. *FASEB J.* **20**, 811–827. (doi:10.1096/fj.05-5424rev)

25. Boulenger V, Nazir TA. 2010 Interwoven functionality of the brain's action and language systems. *Ment. Lex.* **5**, 231–254. (doi:10.1075/ml.5.2.05bou)

26. Pickering MJ, Garrod S. 2013 Forward models and their implications for production, comprehension, and dialogue. *Behav. Brain Sci.* **36**, 377–392. (doi:10.1017/S0140525X12003238)

27. Viswanathan N, Kokkinakis K, Williams BT. 2018 Listeners experience linguistic masking release in noise-vocoded speech-in-speech recognition. *J. Speech, Lang. Hear. Res.* **61**, 428–435. (doi:10.1044/2017_JSLHR-H-17-0215)

28. Viswanathan N, Kokkinakis K. 2019 Listening benefits in speech-in-speech recognition are altered under reverberant conditions. *J. Acoust. Soc. Am.* **145**, EL348–EL353. (doi:10.1121/1.5100898)

29. Ferreira F, Anes MD, Horine MD. 1996 Exploring the use of prosody during language comprehension using the auditory moving window technique. *J. Psycholinguist. Res.* **25**, 273–290. (doi:10.1007/BF01708574)

30. Unsworth N, Robison MK. 2016 Pupillary correlates of lapses of sustained attention. *Cogn. Affect. Behav. Neurosci.* **16**, 601–615. (doi:10.3758/s13415-016-0417-4)

31. Piquado T, Benichov JI, Brownell H, Wingfield A. 2012 The hidden effect of hearing acuity on speech recall, and compensatory effects of self-paced listening. *Int. J. Audiol.* **51**, 576–583. (doi:10.3109/14992027.2012.684403)

32. Mandelbrot BB. 1999 *Multifractals and 1/f Noise*. New York, NY: Springer-Verlag.

33. Ihlen E. 2012 Introduction to multifractal detrended fluctuation analysis in Matlab. *Front. Physiol.* **3**, 141. (doi:10.3389/fphys.2012.00)

34. Mandic DP, Chen M, Gautama T, Van Hulle MM, Constantinides A. 2008 On the characterization of the deterministic/stochastic and linear/nonlinear nature of time series. *Proc. R. Soc. A Math. Phys. Eng. Sci.* **464**, 1141–1160. (doi:10.1098/rspa.2007.0154)

35. Kello CT, Anderson GG, Holden JG, Van Orden GC. 2008 The pervasiveness of 1/f scaling in speech reflects the metastable basis of cognition. *Cogn. Sci.* **32**, 1217–1231. (doi:10.1080/03640210801944898)

36. Van Orden GC, Holden JG, Turvey MT. 2003 Self-organization of cognitive performance. *J. Exp. Psychol. Gen.* **132**, 331–350. (doi:10.1037/0096-3445.132.3.331)

37. Gilden DL. 2001 Cognitive emissions of 1/*f* noise. *Psychol. Rev.* **108**, 33–56. (doi:10.1037/0033-295X.108.1.33)

38. Booth CR, Brown HL, Eason EG, Wallot S, Kelty-Stephen DG. 2018 Expectations on hierarchical scales of discourse: Multifractality predicts both short- and long-range effects of violating gender expectations in text reading. *Discourse Process.* **55**, 12–30. (doi:10.1080/0163853X.2016.1197811)

39. Schreiber T, Schmitz A. 1996 Improved surrogate data for nonlinearity tests. *Phys. Rev. Lett.* **77**, 635–638. (doi:10.1103/PhysRevLett.77.635)

40. Daube C, Ince RAA, Gross J. 2019 Simple acoustic features can explain phoneme-based predictions of cortical responses to speech. *Curr. Biol.* **29**, 1924-1937.e9. (doi:10.1016/j.cub.2019.04.067)

41. Berezutskaya J, Freudenburg Z V, Güçlü U, van Gerven MAJ, Ramsey NF. 2020 Brain-optimized extraction of complex sound features that drive continuous auditory perception. *PLOS Comput. Biol.* **16**, e1007992. (doi:10.1371/journal.pcbi.1007992)

42. Skipper JI, Devlin JT, Lametti DR. 2017 The hearing ear is always found close to the speaking tongue: Review of the role of the motor system in speech perception. *Brain Lang.* **164**, 77–105. (doi:https://doi.org/10.1016/j.bandl.2016.10.004)

43. Fusaroli R, Rączaszek-Leonardi J, Tylén K. 2014 Dialog as interpersonal synergy. *New Ideas Psychol.* **32**, 147–157. (doi:10.1016/j.newideapsych.2013.03.005)

44. Kello CT, Bella SD, Médé B, Balasubramaniam R. 2017 Hierarchical temporal structure in music, speech and animal vocalizations: Jazz is like a conversation, humpbacks sing like hermit thrushes. *J. R. Soc. Interface* **14**, 20170231. (doi:10.1098/rsif.2017.0231)

45. Roeske TC, Kelty-Stephen D, Wallot S. 2018 Multifractal analysis reveals music-like dynamic structure in songbird rhythms. *Sci. Rep.* **8**, 4570. (doi:10.1038/s41598-018-22933-2)

46. Viswanathan N, Kelty-Stephen DG. 2018 Comparing speech and nonspeech context effects across timescales in coarticulatory contexts. *Attention, Perception, Psychophys.* **80**, 316–324. (doi:10.3758/s13414-017-1449-8)

47. Teng DW, Eddy CL, Kelty-Stephen DG. 2016 Non-visually-guided distance perception depends on matching torso fluctuations between training and test. *Attention, Perception, Psychophys.* **78**, 2320–2328. (doi:10.3758/s13414-016-1213-5)

48. Doyon JK, Hajnal A, Surber T, Clark JD, Kelty-Stephen DG. 2019 Multifractality of posture modulates multisensory perception of stand-on-ability. *PLoS One* **14**, e0212220. (doi:10.1371/journal.pone.0212220)

49. Jacobson N, Berleman-Paul Q, Mangalam M, Kelty-Stephen DG, Ralston C. 2021 Multifractality in postural sway supports quiet eye training in aiming tasks: A study of golf putting. *Hum. Mov. Sci.* **76**, 102752. (doi:10.1016/j.humov.2020.102752)

50. Bell C, Carver N, Zbaracki J, Kelty-Stephen D. 2019 Nonlinear amplification of variability through interaction across scales supports greater accuracy in manual aiming: Evidence from a multifractal analysis with comparisons to linear surrogates in the Fitts task. *Front. Physiol.* **10**, 998. (doi:10.3389/fphys.2019.00998)

51. Mangalam M, Carver NS, Kelty-Stephen DG. 2020 Multifractal signatures of perceptual processing on anatomical sleeves of the human body. *J. R. Soc. Interface* **17**, 20200328. (doi:10.1098/rsif.2020.0328)

52. Mangalam M, Carver NS, Kelty-Stephen DG. 2020 Global broadcasting of local fractal fluctuations in a bodywide distributed system supports perception via effortful touch. *Chaos, Solitons & Fractals* **135**, 109740. (doi:10.1016/j.chaos.2020.109740)

53. Carver NS, Bojovic D, Kelty-Stephen DG. 2017 Multifractal foundations of visually-guided aiming and adaptation to prismatic perturbation. *Hum. Mov. Sci.* **55**, 61–72. (doi:10.1016/j.humov.2017.07.005)

54. Stephen DG, Dixon JA. 2011 Strong anticipation: Multifractal cascade dynamics modulate scaling in synchronization behaviors. *Chaos, Solitons & Fractals* **44**, 160–168. (doi:10.1016/j.chaos.2011.01.005)

55. Carver NS, Kelty-Stephen DG. 2017 Multifractality in individual honeybee behavior hints at colony-specific social cascades: Reanalysis of radio-frequency identification data from five different colonies. *Phys. Rev. E* **95**, 22402. (doi:10.1103/PhysRevE.95.022402)

56. Abney DH, Paxton A, Dale R, Kello CT. 2014 Complexity matching in dyadic conversation. *J. Exp. Psychol. Gen.* **143**, 2304–2315. (doi:10.1037/xge0000021)

57. Ramirez-Aristizabal AG, Médé B, Kello CT. 2018 Complexity matching in speech: Effects of speaking rate and naturalness. *Chaos, Solitons & Fractals* **111**, 175–179. (doi:10.1016/j.chaos.2018.04.021)

58. Abney DH, Kello CT, Warlaumont AS. 2015 Production and convergence of multiscale clustering in speech. *Ecol. Psychol.* **27**, 222–235. (doi:10.1080/10407413.2015.1068653)

59. Schloesser DS, Kello CT, Marmelat V. 2019 Complexity matching and coordination in individual and dyadic performance. *Hum. Mov. Sci.* **66**, 258–272. (doi:10.1016/j.humov.2019.04.010)

60. Wood SG, Moxley JH, Tighe EL, Wagner RK. 2017 Does use of text-to-speech and related read-aloud tools improve reading comprehension for students with reading disabilities? A meta-analysis. *J. Learn. Disabil.* **51**, 73–84. (doi:10.1177/0022219416688170)

61. Giannouli V, Banou M. 2020 The intelligibility and comprehension of synthetic *versus* natural speech in dyslexic students. *Disabil. Rehabil. Assist. Technol.* **15**, 898–907. (doi:10.1080/17483107.2019.1629111)

62. Wagner RK, Zirps FA, Edwards AA, Wood SG, Joyner RE, Becker BJ, Liu G, Beal B. 2020 The prevalence of dyslexia: A new approach to its estimation. *J. Learn. Disabil.* **53**, 354–365. (doi:10.1177/0022219420920377)

63. Zablotsky B, Black LI, Maenner MJ, Schieve LA, Danielson ML, Bitsko RH, Blumberg SJ, Kogan MD, Boyle CA. 2019 Prevalence and trends of developmental disabilities among children in the united states: 2009–2017. *Pediatrics* **144**, e20190811. (doi:10.1542/peds.2019-0811)

64. Brysbaert M, New B. 2009 Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behav. Res. Methods* **41**, 977–990. (doi:10.3758/BRM.41.4.977)

65. Marian V, Bartolotti J, Chabal S, Shook A. 2012 CLEARPOND: Cross-Linguistic Easy-Access Resource for Phonological and Orthographic Neighborhood Densities. *PLoS One* **7**, e43230. (doi:10.1371/journal.pone.0043230)

66. Kelty-Stephen D, Lane E, Mangalam M. 2021 Multifractal test for nonlinear changes in time series. *arXiv Prepr. arXiv2105.13113*

67. Chhabra A, Jensen R V. 1989 Direct determination of the $f(\alpha)$ singularity spectrum. *Phys. Rev. Lett.* **62**, 1327–1330. (doi:10.1103/PhysRevLett.62.1327)

68. Halsey TC, Jensen MH, Kadanoff LP, Procaccia I, Shraiman BI. 1986 Fractal measures and their singularities: The characterization of strange sets. *Phys. Rev. A* **33**, 1141–1151. (doi:10.1103/PhysRevA.33.1141)

69. Kelty-Stephen DG, Palatinus K, Saltzman E, Dixon JA. 2013 A tutorial on multifractality, cascades, and interactivity for empirical time series in ecological science. *Ecol. Psychol.* **25**, 1–62. (doi:10.1080/10407413.2013.753804)

70. Bates D, Sarkar D, Bates M, Matrix L. 2007 The lme4 package.

71. R Core Team. 2013 R: A language and environment for statistical computing. (doi:URL https://www.R-project.org/.)

72. Wallot S, O'Brien BA, Haussmann A, Kloos H, Lyby MS. 2014 The role of reading time complexity and reading speed in text comprehension. *J. Exp. Psychol. Learn. Mem. Cogn.* **40**, 1745–1765. (doi:10.1037/xlm0000030)

73. Baars JE, Looman CWN, Steyerberg EW, Kuipers EJ, van der Woude CJ. 2011 Individualised surveillance strategies for colorectal cancer in inflammatory bowel disease. *Gut* **60**, 739–739. (doi:10.1136/gut.2010.229351)

74. Cole SD, Stefanovski D, Towl S, Boyle AG. 2019 Factors associated with prolonged treatment days, increased veterinary visits and complications in horses with subsolar abscesses. *Vet. Rec.* **184**, 251. (doi:10.1136/vr.104138)

75. Spohr SA, Taxman FS, Walters ST. 2015 The relationship between electronic goal reminders and subsequent drug use and treatment initiation in a criminal justice setting. *Addict. Behav.* **51**, 51–56. (doi:10.1016/j.addbeh.2015.07.005)

76. Krebs BL, Torres E, Chesney C, Kantoniemi Moon V, Watters J V. 2017 Applying behavioral conditioning to identify anticipatory behaviors. *J. Appl. Anim. Welf. Sci.* **20**, 155–175. (doi:10.1080/10888705.2017.1283225)

77. Kinsbergen S, Tolsma J, Ruiter S. 2011 Bringing the beneficiary closer: Explanations for volunteering time in dutch private development initiatives. *Nonprofit Volunt. Sect. Q.* **42**, 59–83. (doi:10.1177/0899764011431610)

78. Bone EK, Bouck EC. 2017 Accessible text-to-speech options for students who struggle with reading. *Prev. Sch. Fail. Altern. Educ. Child. Youth* **61**, 48–55. (doi:10.1080/1045988X.2016.1188366)

79. Efron B, Tibshirani R. 1986 Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Stat. Sci.* **1**, 54–75.

80. Shockley K, Baker AA, Richardson MJ, Fowler CA. 2007 Articulatory constraints on interpersonal postural coordination. *J. Exp. Psychol. Hum. Percept. Perform.* **33**, 201–208. (doi:10.1037/0096-1523.33.1.201)

81. Shockley K, Santana M-V, Fowler CA. 2003 Mutual interpersonal postural constraints are involved in cooperative conversation. *J. Exp. Psychol. Hum. Percept. Perform.* **29**, 326–332. (doi:10.1037/0096-1523.29.2.326)

82. Palatinus Z, Kelty-Stephen DG, Kinsella-Shaw J, Carello C, Turvey MT. 2014 Haptic perceptual intent in quiet standing affects multifractal scaling of postural fluctuations. *J. Exp. Psychol. Hum. Percept. Perform.* **40**, 1808–1818. (doi:10.1037/a0037247)

83. Liu H, Bates E, Powell T, Wulfeck B. 1997 Single-word shadowing and the study of lexical access. *Appl. Psycholinguist.* **18**, 157– 180. (doi:10.1017/S0142716400009954)

Table 1. Coefficients of the Poisson regression of self-paced word RT[a] (section 3.1).

effect[b]	***mean***	***s.e.m.*** **(bootstrapped)**	***p***[c] **(bootstrapped)**
intercept	5.54×10^{0}	2.53×10^{-1}	$\mathbf{6.22\times10^{-107}}$
lnSL	3.44×10^{-1}	2.42×10^{-2}	$\mathbf{2.83\times10^{-46}}$
logWF	-3.22×10^{-2}	2.60×10^{-3}	$\mathbf{2.49\times10^{-35}}$
PhonN	-3.97×10^{-4}	2.06×10^{-4}	$\mathbf{2.73\times10^{-02}}$
OrthN	-8.26×10^{-5}	4.88×10^{-4}	4.33×10^{-1}
TTS	-7.16×10^{-2}	1.75×10^{-2}	$\mathbf{2.19\times10^{-5}}$
t_{MF}	-2.77×10^{-5}	1.06×10^{-5}	$\mathbf{4.60\times10^{-3}}$
TTS×logWF	1.65×10^{-2}	3.65×10^{-3}	$\mathbf{3.29\times10^{-6}}$
TTS×PhonN	5.66×10^{-5}	3.48×10^{-4}	4.35×10^{-1}
TTS×OrthN	7.42×10^{-4}	7.53×10^{-4}	1.62×10^{-1}
TTS×t_{MF}	3.46×10^{-5}	1.66×10^{-5}	$\mathbf{1.87\times10^{-2}}$
word#	-2.66×10^{-4}	5.44×10^{-6}	0.00×10^{0}
TTS×word	1.42×10^{-4}	8.04×10^{-6}	$\mathbf{2.55\times10^{-70}}$
1/word#	1.02×10^{0}	3.20×10^{-1}	$\mathbf{7.21\times10^{-4}}$
1/word#×logWF	2.84×10^{-1}	2.60×10^{-3}	$\mathbf{2.49\times10^{-35}}$
1/word#×PhonN	1.73×10^{-2}	1.96×10^{-2}	1.89×10^{-1}
1/word#×OrthN	-8.35×10^{-2}	5.83×10^{-2}	7.61×10^{-2}
1/word#×TTS	-4.06×10^{-1}	5.29×10^{-1}	2.21×10^{-1}
1/word#×t_{MF}	-1.43×10^{-4}	5.21×10^{-4}	3.92×10^{-1}
1/word#×TTS×logWF	-8.41×10^{-2}	1.97×10^{-1}	3.35×10^{-1}
1/word#×TTS×PhonN	8.21×10^{-2}	2.59×10^{-2}	$\mathbf{7.65\times10^{-4}}$
1/word#×TTS×OrthN	3.03×10^{-2}	8.40×10^{-2}	3.59×10^{-1}
1/word#×TTS×t_{MF}	1.81×10^{-3}	6.47×10^{-4}	$\mathbf{2.55\times10^{-3}}$

[a]model in *R* syntax: glmer(I(10*RT) ~ lnSL + I(1/word#)*logWF*condition + I(1/word#)*PhonN*condition + I(1/word#)*OrthN*condition + I(1/word#)*condition*t_{MF} + word#*condition + (1|participant), data = DATA, family = poisson).

[b]abbreviations: lnSL = ln(sound length); logWF = $\log_{10}$(word frequency); PhonN = number of phonological neighbors; OrthN = number of orthographic neighbors; TTS = text-to-speech; t_{MF} = multifractal nonlinearity; word# = word number in the presented sequence.

[c]boldfaced values indicate statistical significance at the alpha level of 0.05.

Table 2. Coefficients of the Poisson regression of cumulative correct answers in the post-narrative test[a] (section 3.2).

effect[b]	***Mean***	***s.e.m.* (bootstrapped)**	***p***[c] **(bootstrapped)**
intercept	1.10×10^{0}	8.39×10^{-1}	9.43×10^{-2}
poly(question,3)1	8.80×10^{0}	1.39×10^{0}	$\mathbf{1.30\times10^{-10}}$
poly(question,3)2	-1.23×10^{0}	1.13×10^{0}	1.37×10^{-1}
poly(question,3)3	1.30×10^{-1}	9.20×10^{-1}	4.44×10^{-1}
typeWRM	2.99×10^{-1}	1.57×10^{-1}	$\mathbf{2.83\times10^{-2}}$
conditionTTS	4.26×10^{0}	1.44×10^{0}	$\mathbf{1.52\times10^{-3}}$
log(RT-mean)	2.46×10^{0}	4.90×10^{-1}	$\mathbf{2.46\times10^{-7}}$
log(RT-median)	-2.48×10^{0}	5.59×10^{-1}	$\mathbf{4.47\times10^{-6}}$
RT-$\Delta\alpha$	4.67×10^{-1}[d]	7.65×10^{-1}	2.71×10^{-1}[d]
RT-t_{MF}	-7.11×10^{-3}[e]	3.20×10^{-3}	$\mathbf{1.32\times10^{-2}}$[e]
poly(question,3)1×typeWRM	9.46×10^{-1}	9.07×10^{-1}	1.48×10^{-1}
poly(question,3)2×typeWRM	-3.17×10^{0}	8.34×10^{-1}	$\mathbf{7.02\times10^{-5}}$
poly(question,3)3×typeWRM	1.16×10^{0}	6.32×10^{-1}	$\mathbf{3.33\times10^{-2}}$
conditionTTS×log(RT-mean)	-4.12×10^{0}	5.91×10^{-1}	$\mathbf{1.63\times10^{-12}}$
conditionTTS×log(RT-median)	3.58×10^{0}	6.49×10^{-1}	$\mathbf{1.74\times10^{-8}}$
poly(question,3)1×RT-$\Delta\alpha$	16.65×10^{0}	5.95×10^{0}	$\mathbf{2.57\times10^{-3}}$
poly(question,3)2×RT-$\Delta\alpha$	-5.74×10^{0}	4.67×10^{0}	1.10×10^{-1}
poly(question,3)3×RT-$\Delta\alpha$	3.51×10^{0}	3.53×10^{0}	1.60×10^{-1}
typeWRM×conditionTTS	1.99×10^{-2}	2.36×10^{-1}	4.66×10^{-1}
typeWRM×RT-$\Delta\alpha$	-4.66×10^{0}	8.37×10^{-1}	$\mathbf{1.31\times10^{-8}}$
typeWRM×RT-t_{MF}	2.81×10^{-2}	3.57×10^{-3}	$\mathbf{1.62\times10^{-15}}$
conditionTTS×RT-$\Delta\alpha$	-2.28×10^{0}[e]	9.90×10^{-1}	$\mathbf{1.06\times10^{-2}}$[e]
conditionTTS×RT-t_{MF}	6.57×10^{-3}[d]	4.02×10^{-3}	5.12×10^{-2}[d]
typeWRM×conditionTTS×RT-$\Delta\alpha$	2.53×10^{0}	1.23×10^{0}	$\mathbf{2.01\times10^{-2}}$
typeWRM×conditionTTS×RT-t_{MF}	-3.13×10^{-2}	4.86×10^{-3}	$\mathbf{6.36\times10^{-11}}$

[a]model in *R* syntax: glmer(cum-corr ~ + poly(question,3) + poly(question,3)*type + condition*[log(RT_mean) + log(RT_median)] + poly(question,3)*RT_$\Delta\alpha$ + type*condition*RT_$\Delta\alpha$ + type*condition*RT_t_{MF} + (1|participant), data = DATA, family = poisson).

[b]abbreviations: poly(question,3)i = i[th]-order orthogonal polynomial ($1\leq i\leq 3$); LD-status = 1 or 0 depending on whether participant did or did not self-report learning disability; typeWRM = word-recognition memory subset of memory test; conditionTTS = condition involving self-paced listening with text-to-speech; RT = response time; RT-$\Delta\alpha$ = multifractal spectrum width of RT sequence; RT-t_{MF} = multifractal nonlinearity in RT sequence.

[c]boldfaced values indicate statistical significance at the alpha level of 0.05.

[d]significant and reversed sign with the inclusion of LD-status.

[e]no longer significant with the inclusion of LD-status.

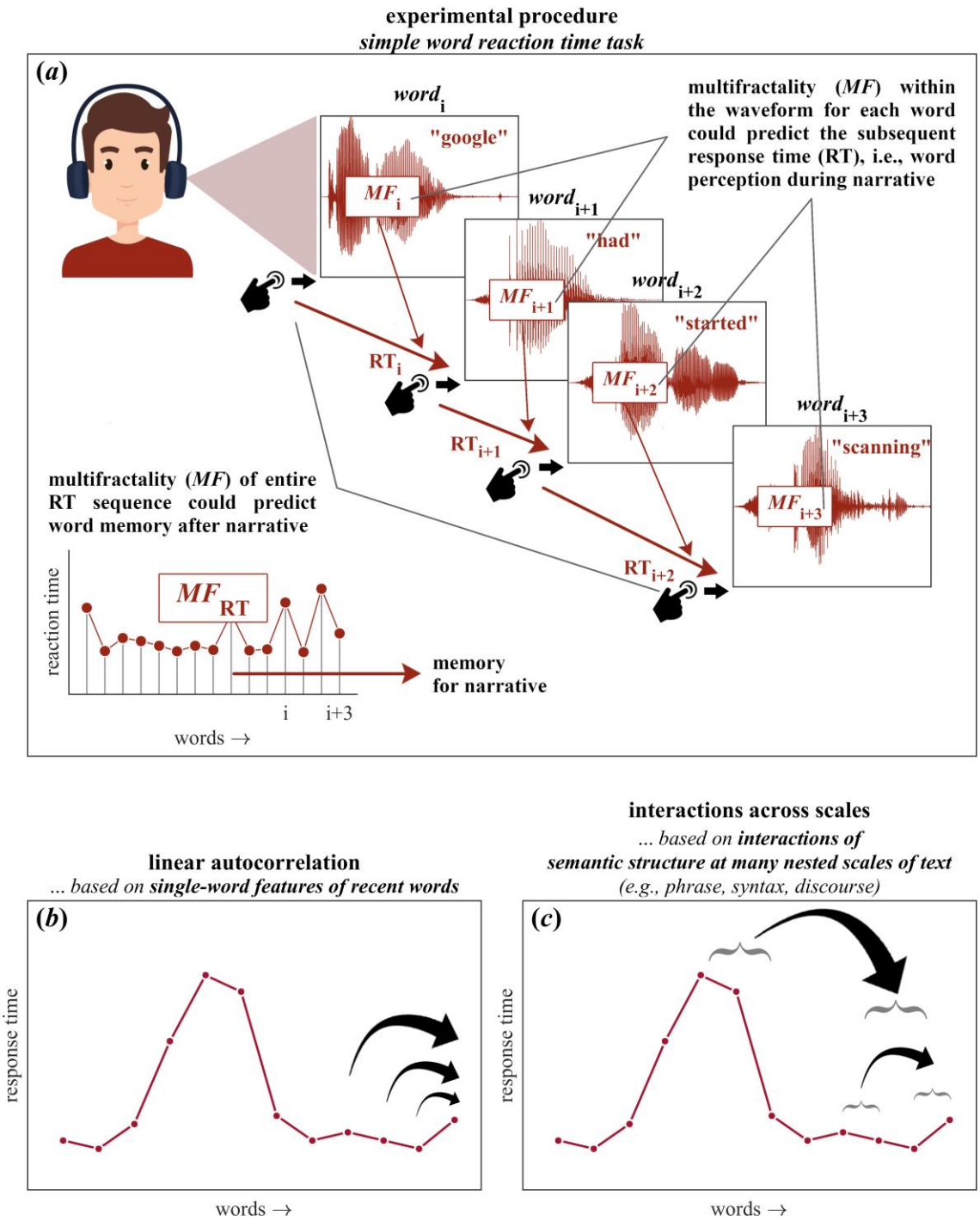


Figure 1. Experimental procedure and two perspectives response times (RT). (*a*) Each word presented to a participant through headphones elicited a button press when the participant was ready for the next word. The multifractal and psycholinguistic properties of each word were used to predict response time to test Hypothesis-1. Multifractal attributes (and standard descriptive statistics) for each Hypothesis-2: whether multifractal attributes of the RT series predicted memory and comprehension of the story (see Section 2.2 for details). (*b*) The autoregressive perspective that each word's RT entails the cumulative summing of subsequent and independent processes regarding lexical aspects of preceding words, with longer arrows indicating effects of previous behavior at independent, non-overlapping time lags. (*c*) The multifractal perspective that each word's RT entails the

interactions unfolding among processes across many nested time scales of text (e.g., word, phrase, and discourse), with arrows indicating the effects of previous behavior at different time lags. Horizontal braces schematize how effects of previous behavior at longer, coarser scales (top of panel c) may influence shorter, finer scales.

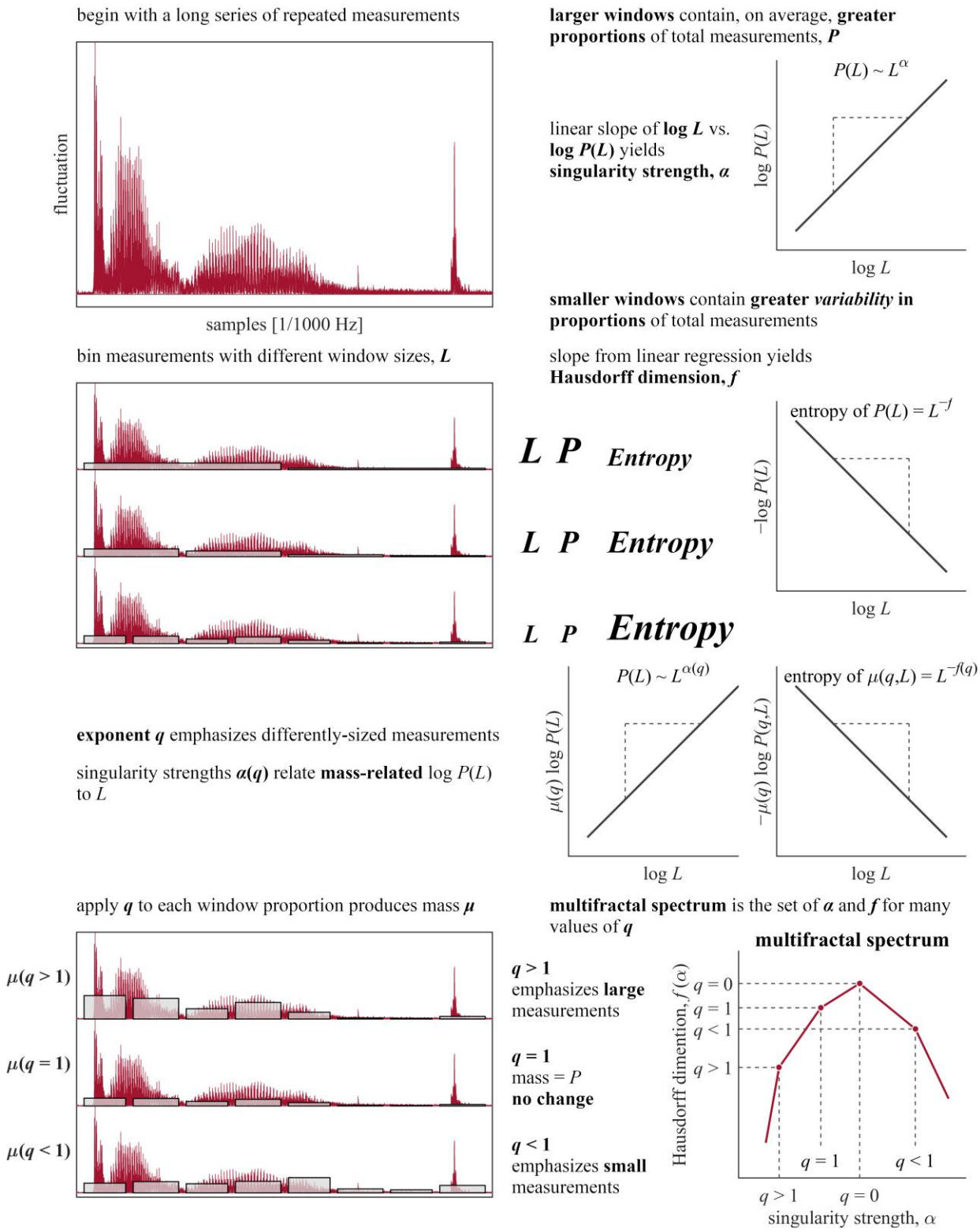


Figure 2. Chhabra and Jensen's [66] method used to estimate multifractal spectrum $f(\alpha)$. The method evaluates how well a measured series (top left) fits a general expectation about relationships between proportion and the length of observation. This procedure tiles the series with nonoverlapping windows of progressively greater length (2nd row left). Windowed subsets will cover greater proportion of total series as length increases (top right), with

lesser Shannon entropy (2nd row right). Smaller and larger values of a continuous q exponent emphasize sparser and denser portions of the series, respectively (3rd row left). Applying q to window proportions produces a "mass" quantity (bottom left). "Mass"-weighted proportion and the Shannon entropy of "mass" both exhibit scaling relationships with window length (3rd row right). Scaling exponents for these relationships at each q value specify an asymmetric inverted-U shaped curve known as the multifractal (or singularity) spectrum. See Section 2.4.1 for further details.

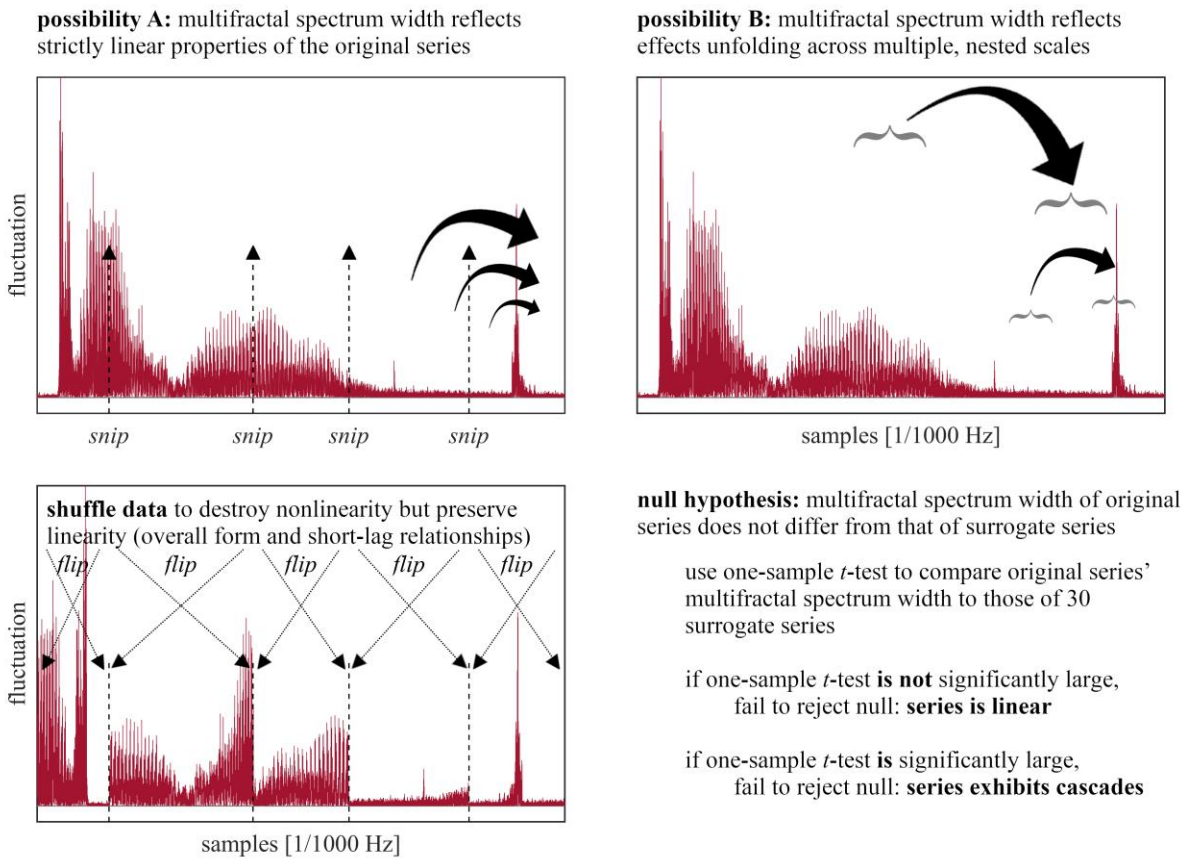


Figure 3. Surrogate analysis used to identify whether a series exhibits multifractality due to nonlinearity (cascades). Speech waveforms may reflect effects unfolding over independent timescales (top left) or interacting timescales (top right). IAAFT procedure preserving linear structure while destroying the original sequence and so any potential interactions across scale. Multifractal-spectrum widths differences between original and surrogate series offers indicates strength of interactions across timescale. See Section 2.4.2 for further details.

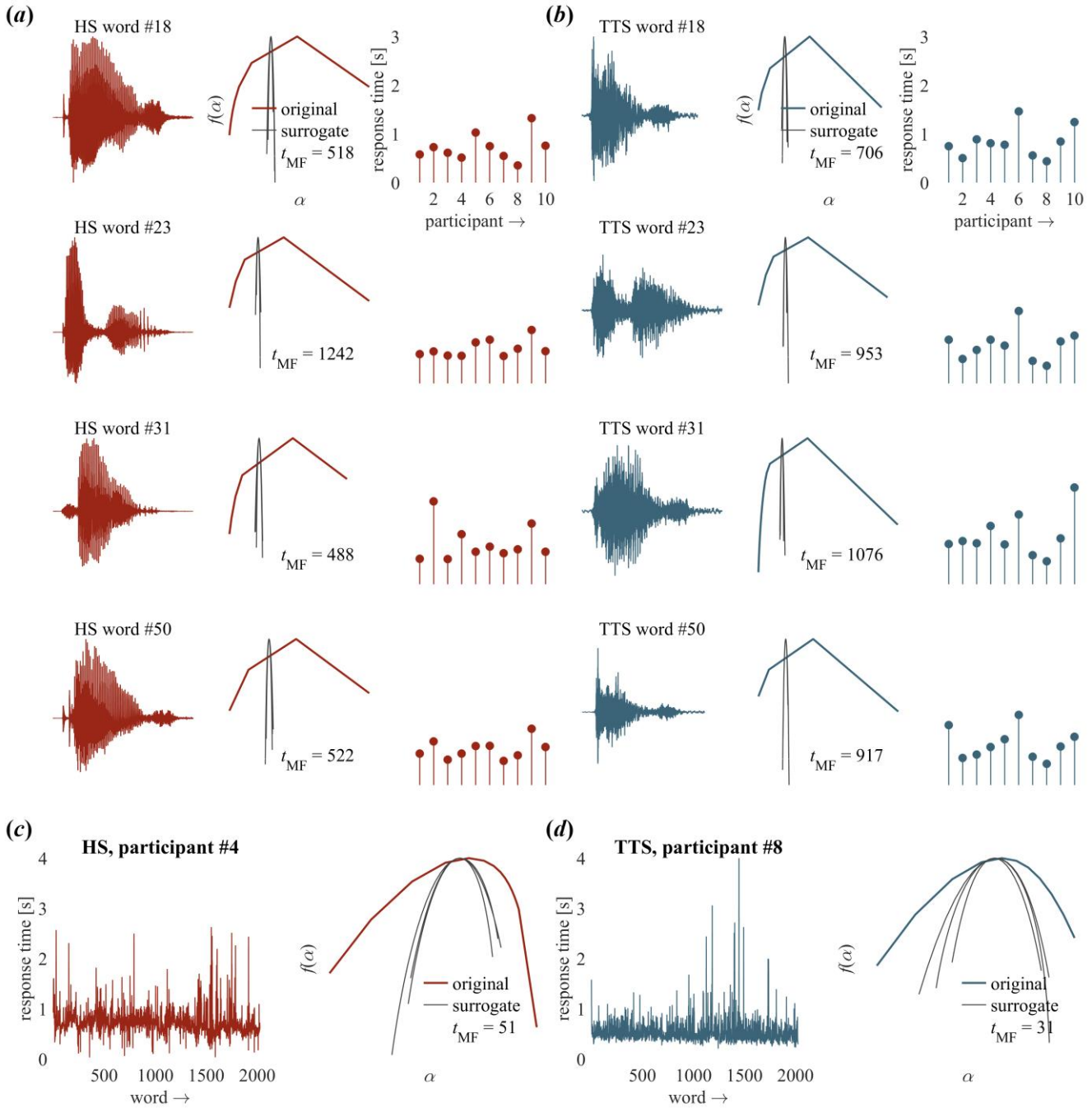


Figure 4. Multifractal spectra of representative stimuli and response times (RTs). Human-speech (a) and text-to-speech (b) words corresponding response times (RTs) for each participant (n = 10 in each condition). Multifractal spectrum of RTs for a representative participant hearing human speech (*c*) and for a representative participant hearing text-to-speech condition (*d*).

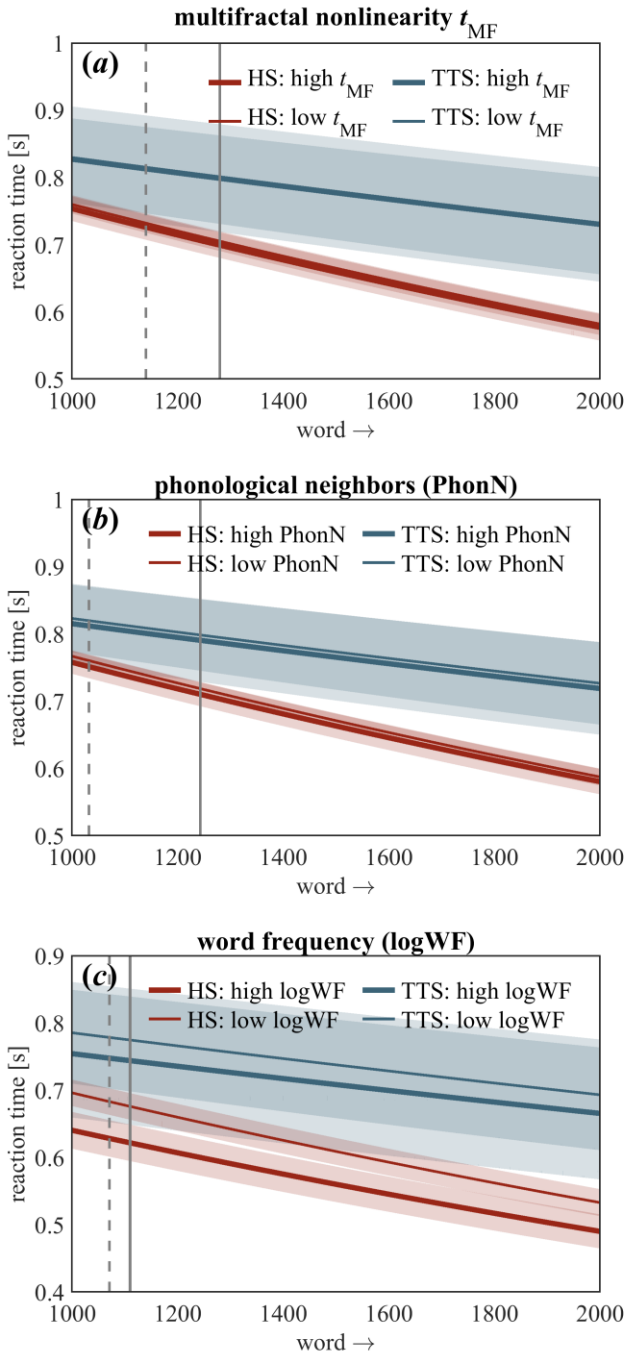


Figure 5. Model predictions for response time (section 3.1.2). Differences between human speech (HS) and text-to-speech (TTS) appear as marked by vertical dashed and solid lines for 1st and 3rd quartiles, respectively, of various predictors: words #1140 and #1280 for t_{MF} (*a*), words #1032 and #1243 for PhonN (*b*), and words #1071 and #1110 for logWF (*c*), respectively. Shaded bars indicate ±2*s.e.m.*

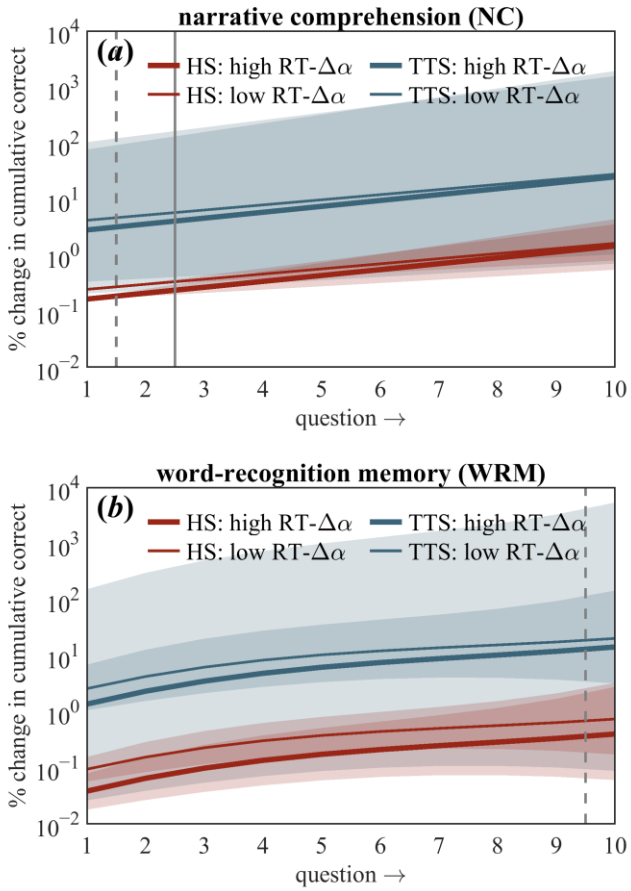


Figure 6. Model predictions for performance in narrative comprehension and word-recognition memory tests (section 3.2). Effect of RT-$\Delta\alpha$ on % change in cumulative correct answers in narrative comprehension test (*a*) and in word-recognition memory test (*b*). Dashed and solid gray vertical lines indicate when low RT-$\Delta\alpha$ and high RT-$\Delta\alpha$, respectively, stops predicting better test performance compared to low RT-$\Delta\alpha$ and high RT-$\Delta\alpha$, respectively, in speech. Shaded bars indicate ± 2*s.e.m.*

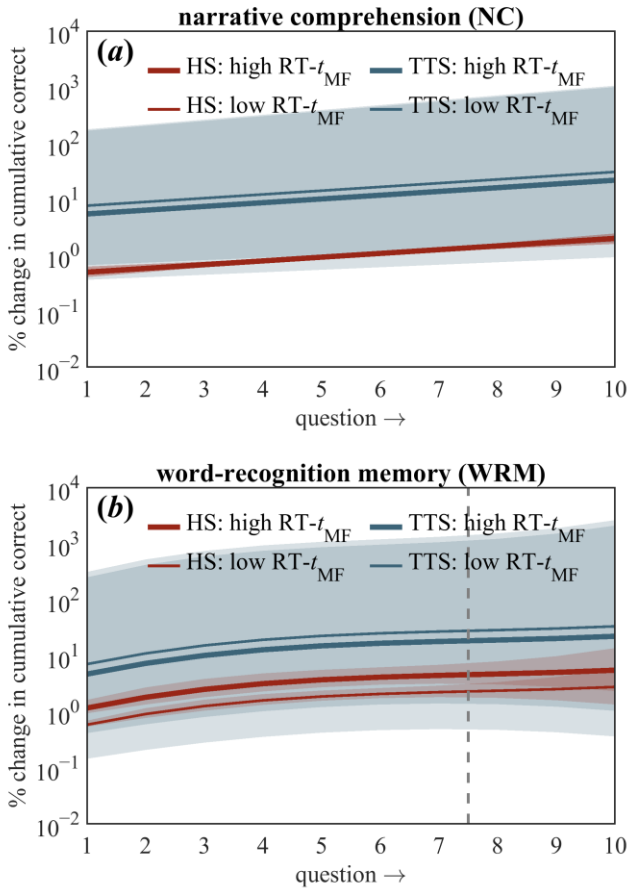
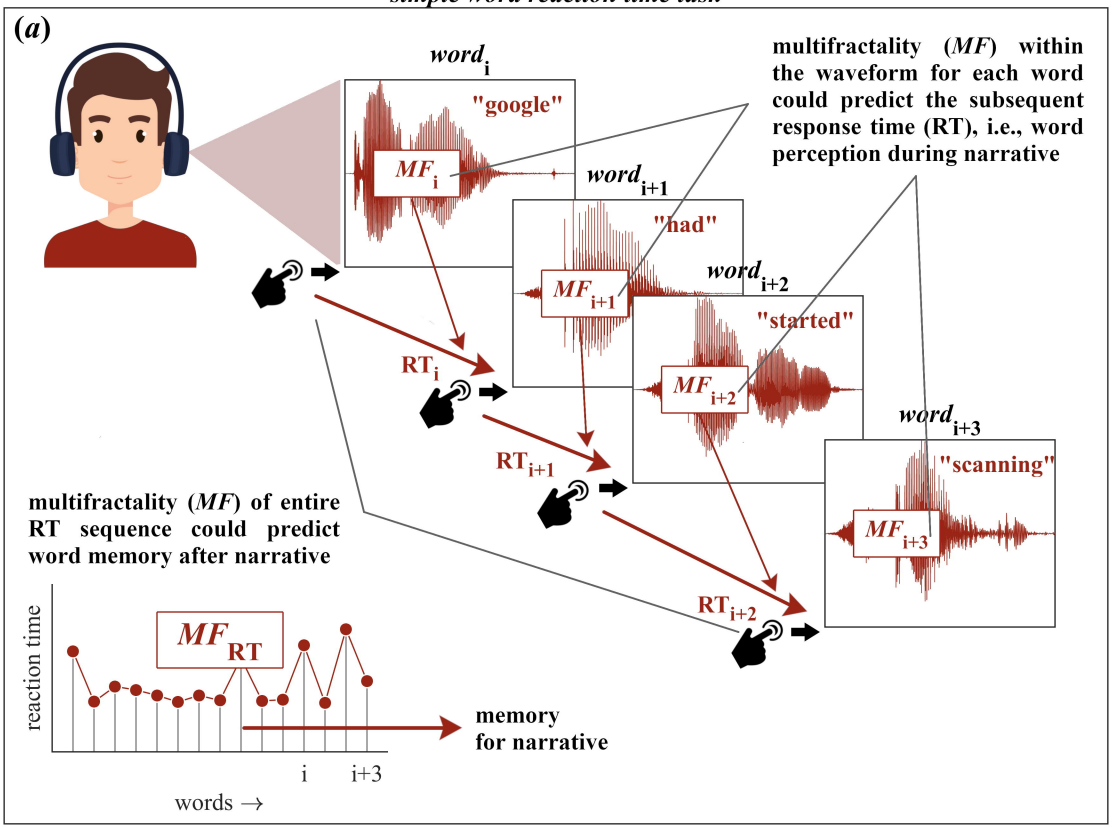


Figure 7. Model predictions for performance in narrative comprehension and word-recognition memory tests (section 3.2). Effect of RT-t_{MF} on % change in cumulative correct answers in narrative comprehension test (*a*) and in word-recognition memory test (*b*). Dashed vertical line indicates when low RT-t_{MF} stops predicting better test performance compared to low RT-t_{MF} in speech. Shaded bars indicate ±2*s.e.m.*

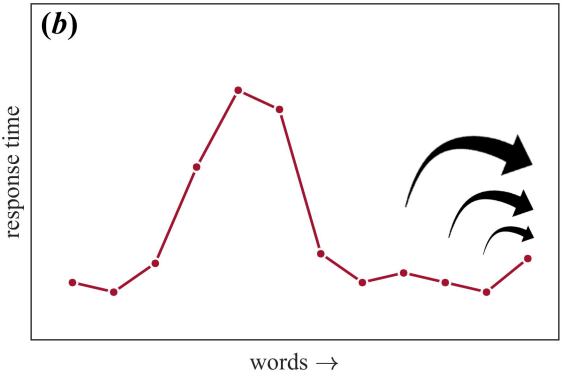
experimental procedure

simple word reaction time task



linear autocorrelation

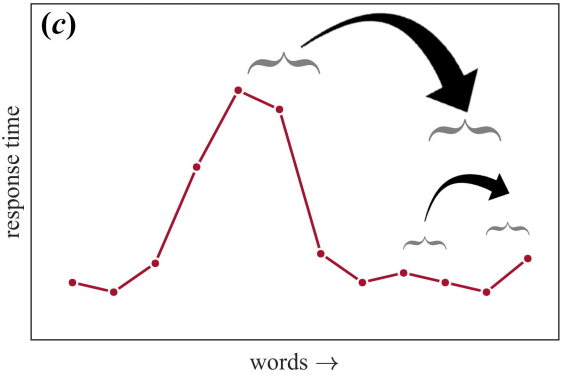
... based on ***single-word features of recent words***



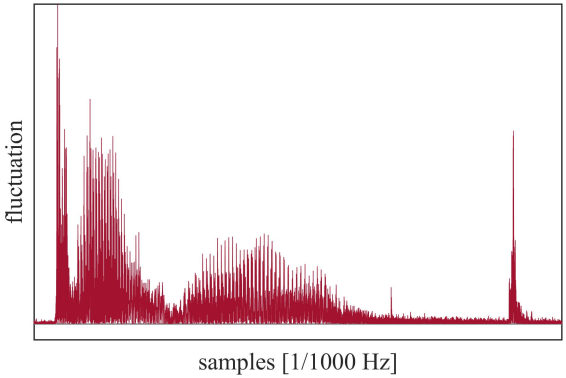
interactions across scales

... based on ***interactions of semantic structure at many nested scales of text***

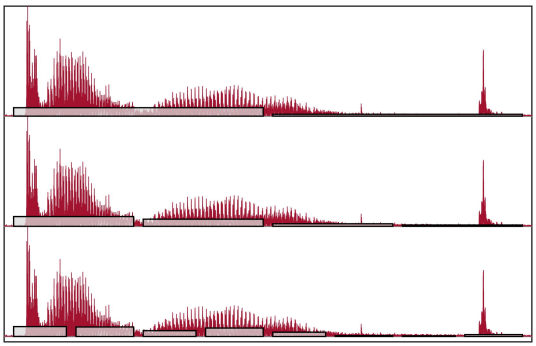
(e.g., phrase, syntax, discourse)



begin with a long series of repeated measurements



bin measurements with different window sizes, ***L***



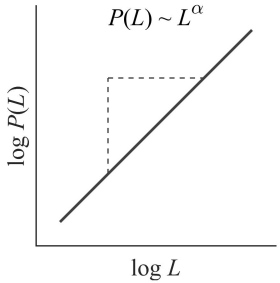
L ***P*** ***Entropy***

L ***P*** ***Entropy***

L ***P*** ***Entropy***

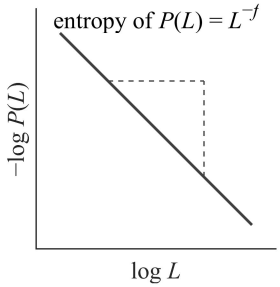
larger windows contain, on average, **greater proportions** of total measurements, ***P***

linear slope of **log *L*** vs. **log *P*(*L*)** yields **singularity strength, *α***



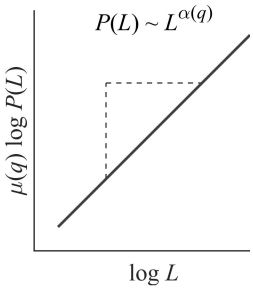
smaller windows contain **greater *variability* in proportions** of total measurements

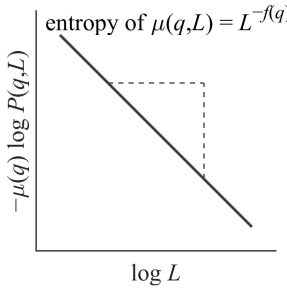
slope from linear regression yields **Hausdorff dimension, *f***



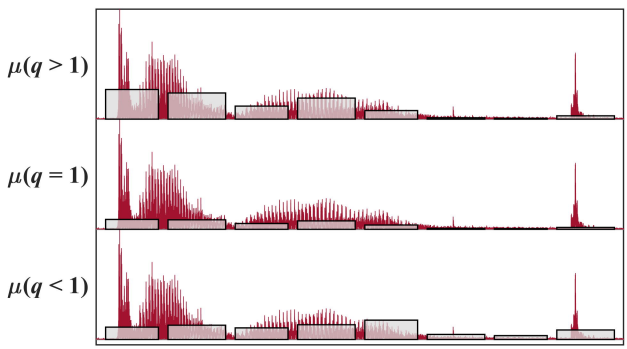
exponent *q* emphasizes differently-sized measurements

singularity strengths ***α*(*q*)** relate **mass-related** log $P(L)$ to L





apply ***q*** to each window proportion produces mass ***μ***

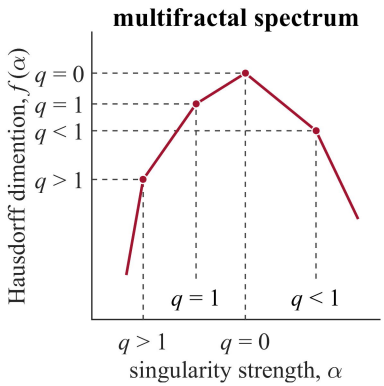


***q* > 1**
emphasizes **large** measurements

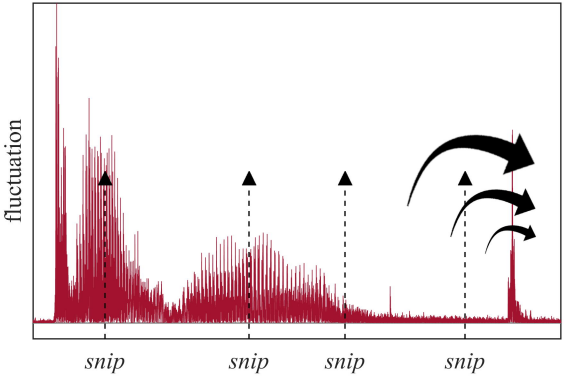
***q* = 1**
mass = P
no change

***q* < 1**
emphasizes **small** measurements

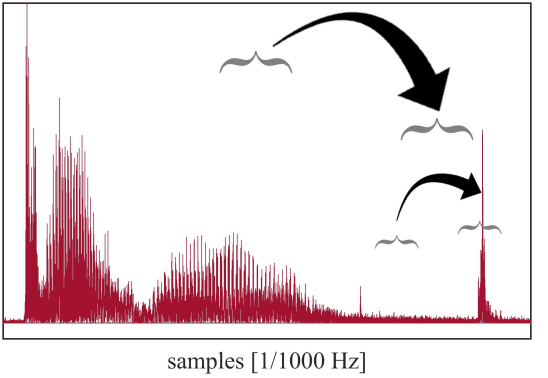
multifractal spectrum is the set of ***α*** and ***f*** for many values of ***q***

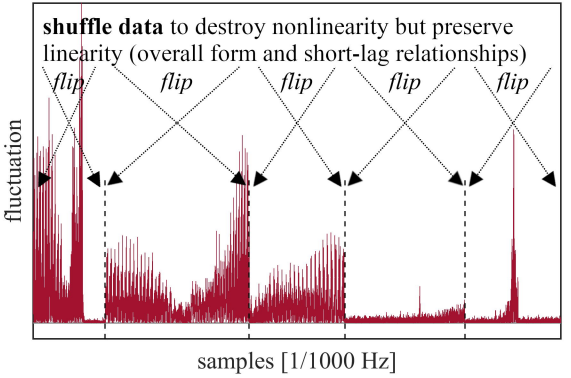


possibility A: multifractal spectrum width reflects strictly linear properties of the original series



possibility B: multifractal spectrum width reflects effects unfolding across multiple, nested scales



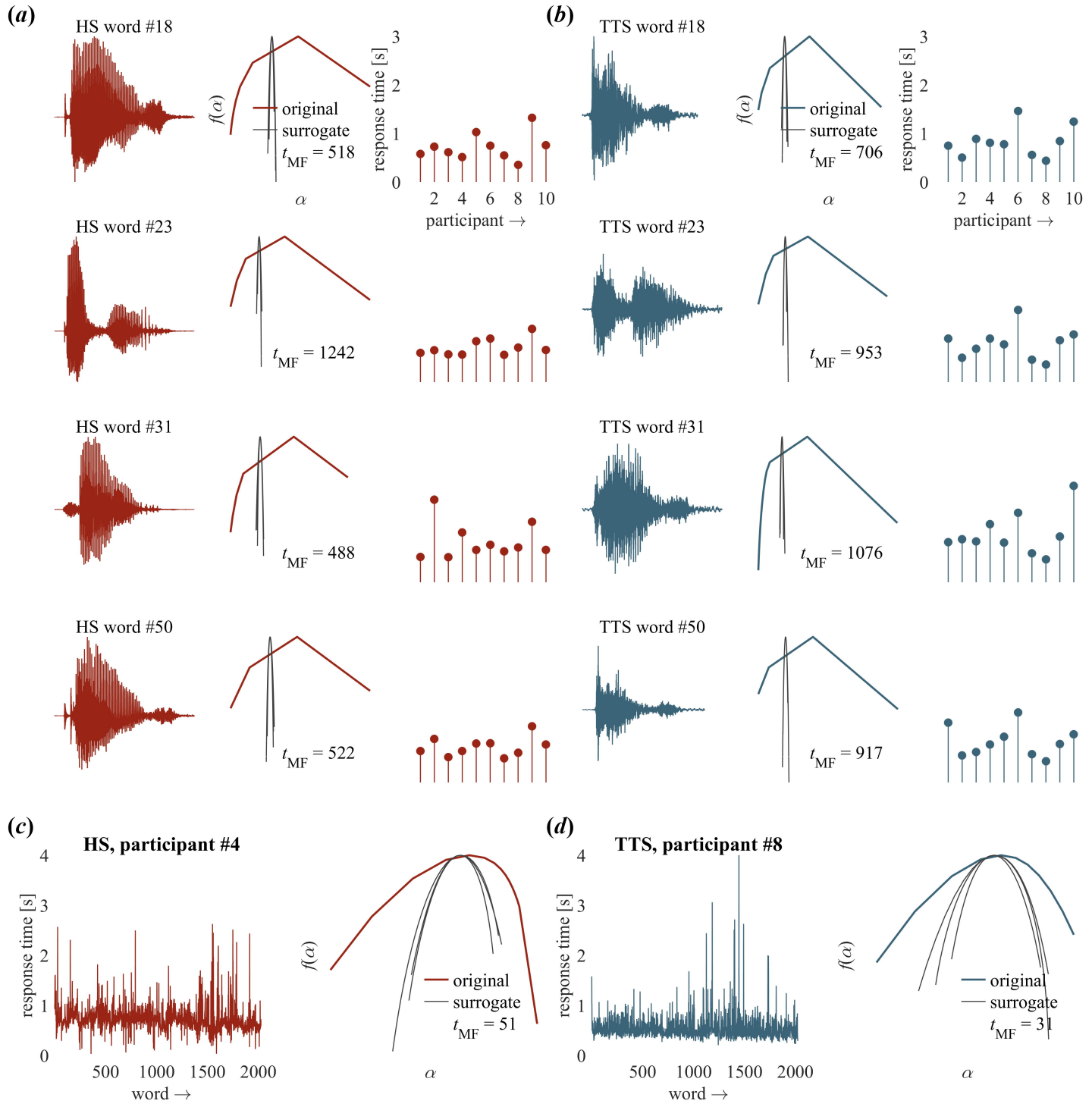


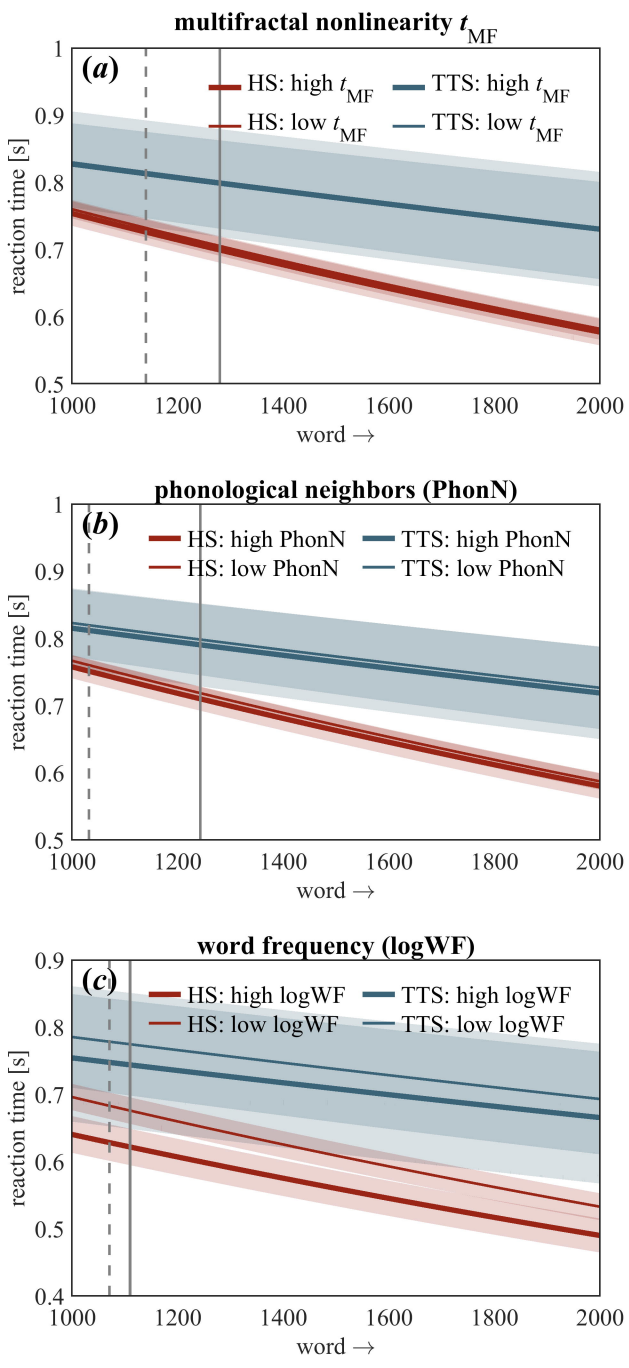
null hypothesis: multifractal spectrum width of original series does not differ from that of surrogate series

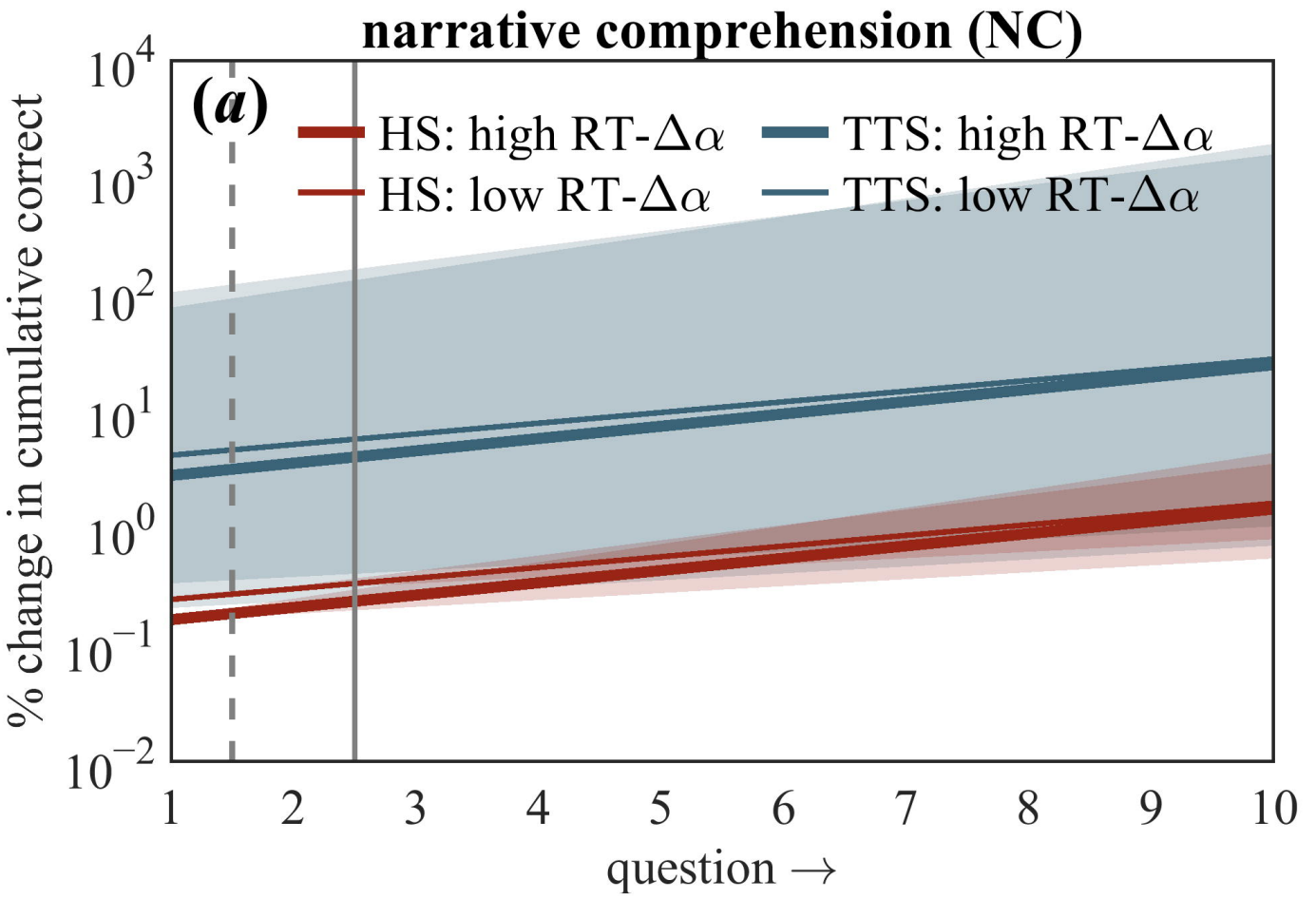
use one-sample *t*-test to compare original series' multifractal spectrum width to those of 30 surrogate series

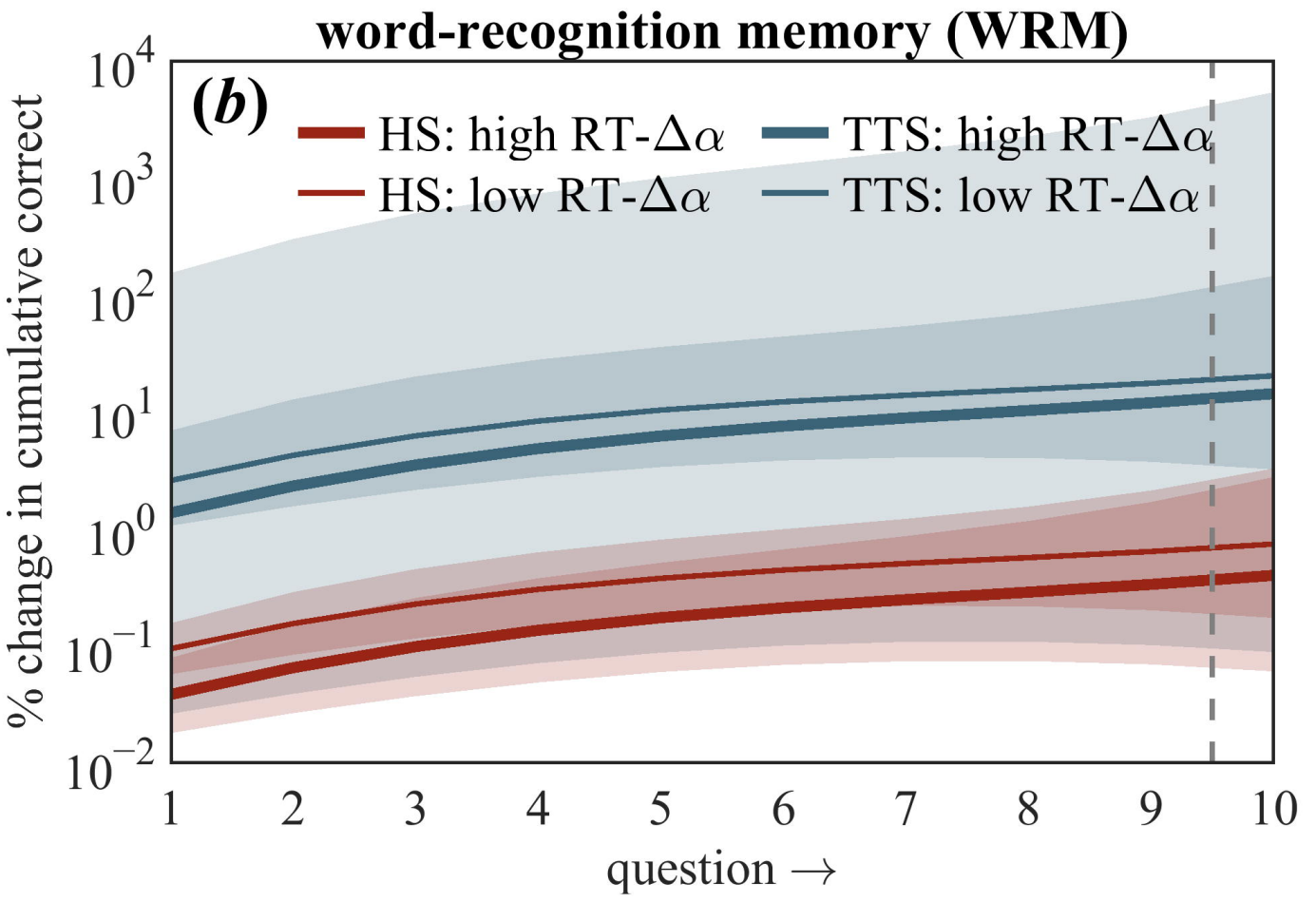
if one-sample *t*-test **is not** significantly large,
fail to reject null: **series is linear**

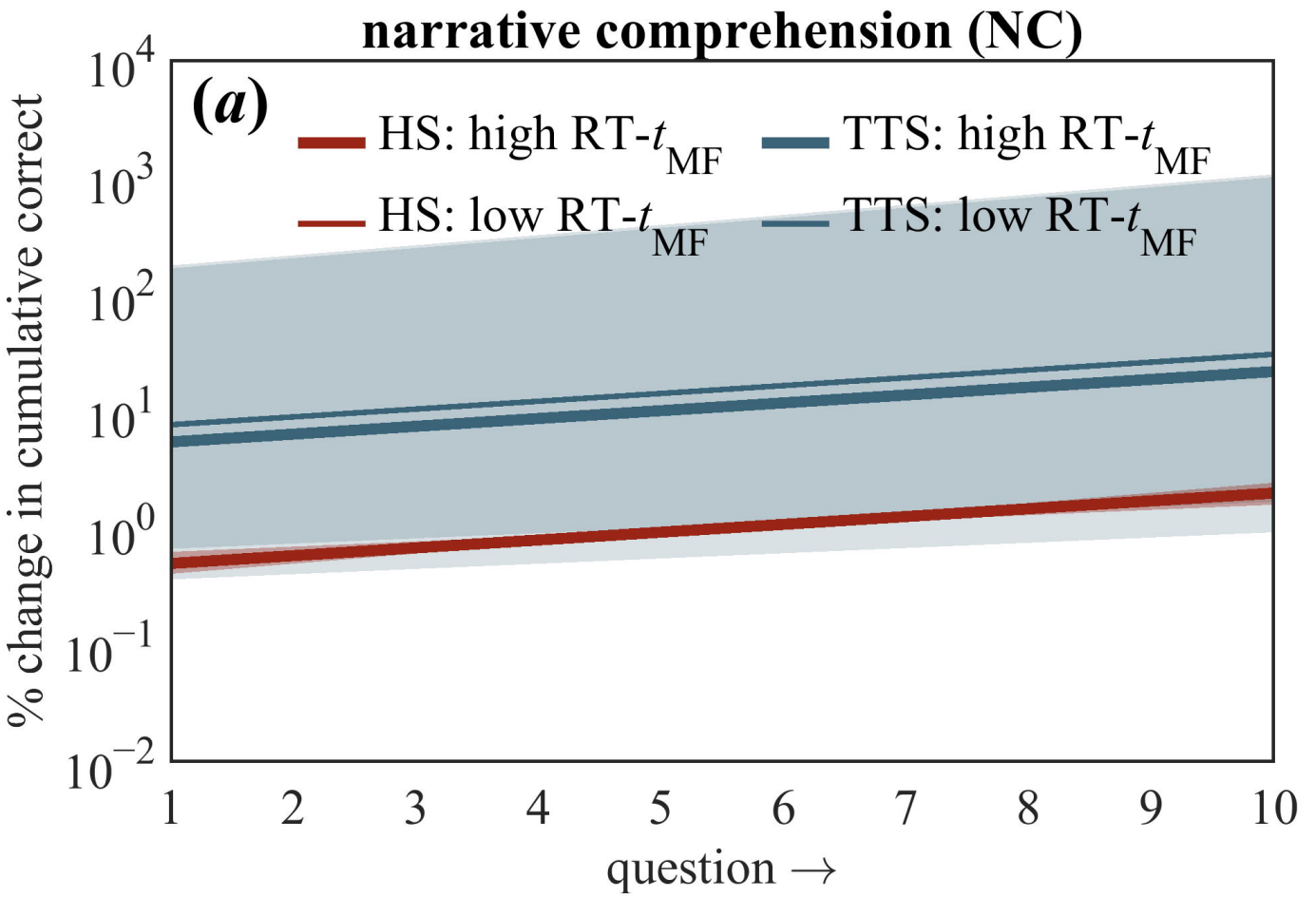
if one-sample *t*-test **is** significantly large,
fail to reject null: **series exhibits cascades**

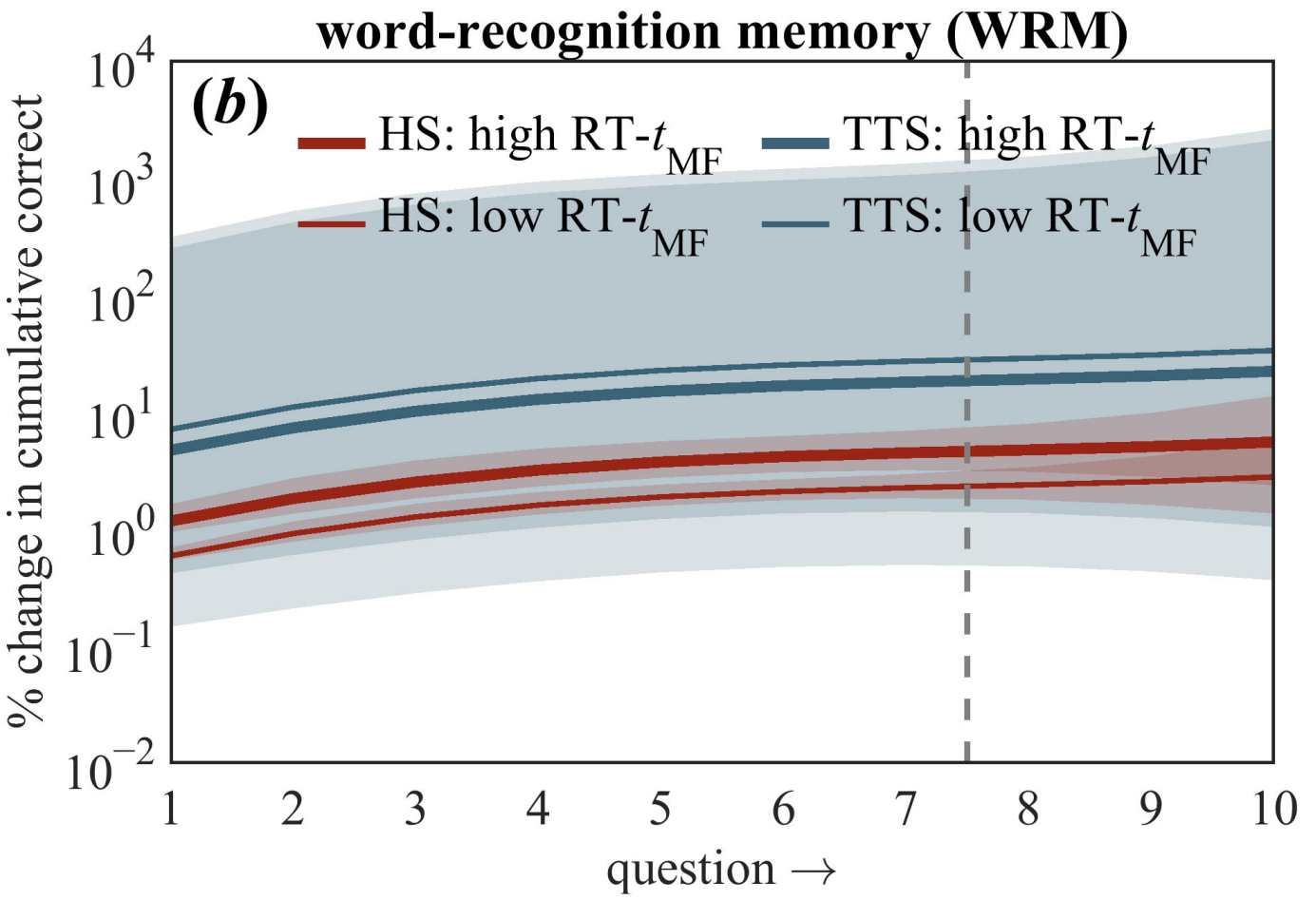

(a)
HS word #18
original
surrogate
t_{MF} = 518
$f(\alpha)$
α
response time [s]
participant →
HS word #23
t_{MF} = 1242
HS word #31
t_{MF} = 488
HS word #50
t_{MF} = 522
(b)
TTS word #18
original
surrogate
t_{MF} = 706
TTS word #23
t_{MF} = 953
TTS word #31
t_{MF} = 1076
TTS word #50
t_{MF} = 917
(c)
HS, participant #4
response time [s]
word →
original
surrogate
t_{MF} = 51
(d)
TTS, participant #8
t_{MF} = 31


multifractal nonlinearity t_{MF}
(a)
HS: high t_{MF}
HS: low t_{MF}
TTS: high t_{MF}
TTS: low t_{MF}
reaction time [s]
word →
phonological neighbors (PhonN)
(b)
HS: high PhonN
HS: low PhonN
TTS: high PhonN
TTS: low PhonN
reaction time [s]
word →
word frequency (logWF)
(c)
HS: high logWF
HS: low logWF
TTS: high logWF
TTS: low logWF
reaction time [s]
word →


narrative comprehension (NC)
(a)
HS: high RT-Δα
HS: low RT-Δα
TTS: high RT-Δα
TTS: low RT-Δα
% change in cumulative correct
question →


word-recognition memory (WRM)
(b)
HS: high RT-Δα
HS: low RT-Δα
TTS: high RT-Δα
TTS: low RT-Δα
% change in cumulative correct
question →


narrative comprehension (NC)
(a)
HS: high RT-t_{MF}
HS: low RT-t_{MF}
TTS: high RT-t_{MF}
TTS: low RT-t_{MF}
% change in cumulative correct
question →


word-recognition memory (WRM)
(b)
HS: high RT-t_{MF}
HS: low RT-t_{MF}
TTS: high RT-t_{MF}
TTS: low RT-t_{MF}
% change in cumulative correct
question →