Fractal and recurrence analysis of psycholinguistic data

Sebastian Wallot¹, Beth A. O'Brien² and Guy Van Orden¹

¹CAP Center for Cognition, Action and Perception, University of Cincinnati / ²Early Childhood Education & Human Development, University of Cincinnati

This chapter illustrates nonlinear methods of data analyses for timeseries of reading times, obtained from a self-paced reading task. We present reading data obtained from adult readers, highlighting differences in reading fluency between literate adults (Wallot & Van Orden, 2011b, c), as well as reading data obtained from 2nd, 4th, and 6th graders, highlighting differences in reading performance between children of different ages. These two different data sets illustrate choices that are encountered when nonlinear analysis techniques of timeseries are employed, and we will discuss solutions to maximize the validity and reliability of the analyses.

Study 1 concerns the adult data which present a three-factor design with two levels of each factor (i.e., $2 \times 2 \times 2$): the size of the text unit — *word units* versus *sentence units*; the fluency of readers — *Ph.D. candidates* in English literature versus *undergraduates* from an introduction to psychology course; and text repetition — *first reading* versus a *subsequent reading*. The nonlinear methods illustrated are timeseries analyses, which give information about the hypothetical system dynamics that could have produced the time-ordered variability observed across reading times. The payoff from these analyses is new information about performance over-and-above and different from the results of conventional analyses.

Study 2 concerns the children's reading data containing only a single factor, grade level (2nd, 4th, and 6th). Nonlinear methods here illustrate the differences between the reading performance of the older and younger children, to describe the acquisition of reading skill.

In both cases, conventional analyses reveal little about the systems under study. The nonlinear analyses, on the other hand, reveal differences due to the text unit conditions and the reading fluency of the adult's performance, as well as differences in reading performance across the groups of children. The results give

a more nuanced picture of what happened in the self-paced reading experiments, and more nuanced pictures are more usefully constraining of theoretical hypotheses. The promise then is a greater capacity to discriminate hypotheses about the reading process.

Two general categories of methods are introduced — fractal analysis and recurrence quantification analysis — but before that we describe the methods of the self-paced reading task and the picture of the data that is created by conventional statistical analyses. We then introduce spectral analysis to illustrate fractal methods, describing the step-by-step application of the analyses. After that we do the same with recurrence quantification methods, ending with a discussion of all the results obtained and how they might speak to reading performance. Throughout, we provide references to in-depth discussions of the techniques.

Self-paced reading by adults and children

In the self-paced reading task, readers advance through a text at their own pace, which they control using the spacebar of a computer keyboard. Each time a participant reader presses the spacebar the next unit of text appears on the computer screen. The time that passes between key presses estimates the reading time of the text unit. The adult reading time data were sampled from a story presented either in word units or sentence units. Thus, text accumulated on the screen either word-by-word or sentence-by-sentence.

The adult readers read a short story titled *The Arelis Complex* by Louis P. DeGrado (2003). This story describes fictive intergalactic politics of the Arelians, who were nearly wiped out in a previous conflict with another civilization. Star ship officer Drakh Norh, charged with the safety of the Arelian home planet, oversees the interception of an unmanned space vessel. The vessel is traced back to a 'blue planet' whose inhabitants had been deemed incapable of such technology. Officer Drakh concludes that the inhabitants of the blue planet may develop eventually into a threat and leads a war fleet to destroy its population. The attempt fails, however, leaving the Arelians facing an uncertain future.

The *Arelis Complex* consists of 13,930 words, which in turn make up 1,042 sentences. Text units were presented in Times New Roman font (13.5 pt.) on a standard computer monitor, paced by the reader's spacebar presses. Participants read the story while seated before the computer monitor, pressing the spacebar to bring text to the screen. Figure 1 illustrates the self-paced reading procedure.

The key factor of the adult reading time study was the reading fluency of participants, manipulated in a quasi-experimental design. The participants were 32, Ph.D. candidate, graduate students of English Literature and 32 freshman

Figure 1. Illustration of the self-paced reading task for sentence-units.

undergraduates attending the University of Cincinnati. We assumed that the graduate students were the more fluent readers, and the data analyses tested for reliable differences in reading performance between graduate student readers and undergraduates. Also, 16 of the graduate students and 16 of the undergraduates read the story more than once.

All adult participants had normal or corrected to normal vision and reported being unfamiliar with the story. Reading the story took between 30 and 80 minutes, depending on the individual. Participants were compensated for their participation in class credit or cash. After reading the story each reader completed an exam, of which they had been forewarned, to assess story comprehension and memory. The exam required a written summary of the story plot, indicating characters and their roles in the story, plus a multiple-choice sentence completion task. Items for the sentence completion task covered the entire story.

The participating children read the story 'Alien and Possum: Hanging Out' by Tony Johnston (2002). The text describes two friends, Alien and Possum, and the various things they do over several days (such as having conversations about their lives, celebrating birthday or climbing on trees).

The difficulty of the text was rated to be appropriate for somewhat advanced 2nd graders (ATOS book level of 2.5), with the content being rated as interesting (4.5 out of 5) by readers (Accelerated Reader, 2011). The original text consisted of 1201 words, but was carefully trimmed down to 1099 words to reduce reading time for the children. The story was presented only word-by-word, so that the obtained ≈ 1100 word reading times satisfied the lower bound of what was

needed to conduct the analyses (see Holden, 2005; Wallot & Van Orden, 2011c). The illustrations of the book were removed as well.

The key factor examined were the reading times of children in different grades. Thirteen children participated in this pilot study, three 2nd graders, five 4th graders, and five 6th graders. All were native speakers of English and all had normal or corrected to normal vision. Also, all reported that they were unfamiliar with the story. Testing took place at the children's elementary school. Reading took 10 to 20 minutes, and participants were rewarded with a little gift. After reading the story, participants were asked 6 multiple-choice questions about the story, of which they had been forewarned, to assess their comprehension. The presentation mode was the same as for the adult participants (see Figure 1) employing only word units.

Study 1: Adults' reading times

Typical data analyses in psycholinguistics estimate parameters such as mean and variance of reading times, thought to correspond to population parameters. Contrasts focus on differences among means (and differences among differences to gauge interaction effects). These are all static contrasts however, existing outside of time, even when measured at two different points in time. The present nonlinear structures in measured values manifest over time. Hence, the analysis of interest must respect the ordered progression of time during the collection of the data. It would matter not all that we randomize the data before computing mean and variance, whereas randomized data lose all information stemming from the order in which, or time at which they were collected.

Figure 2 presents average reading times bracketed by standard error bars in the word and sentence presentation conditions. An analysis of variance included the factors *words* vs. *sentences*, *graduates* vs. *undergraduates*, and *first reading* vs. *re-reading*. We observed reliable interaction effects between text unit and reader group ($F(1, 56) = 6.58, p < .05$) and between text unit and repeated reading ($F(1, 56) = 15.27, p < .001$), plus main effects of text unit ($F(1, 56) = 300.47, p < .001$), reader group ($F(1, 56) = 7.54, p < .01$) and repeated reading ($F(1, 56) = 16.48, p < .001$). Words units are read faster than sentence units, graduate students read faster on average than undergraduate students, and a repeated reading of the text yielded faster average reading times than the first reading.

Separate analyses of word unit and sentence unit conditions shed light on the previous interaction effects: When the story appeared word by word, graduate and undergraduate readers produced average reading times that did not differ statistically. There were no reliable differences in average reading times between undergraduate and graduate readers, nor between repeated readings of the text

Figure 2. Top panels: Average and standard deviation of word reading times in seconds bracketed by standard error. Bottom panels: Average and standard deviation of sentence reading times in seconds bracketed by standard error.

presented word by word ($F < 2.08$). However, when the text was presented sentence by sentence, undergraduate readers were slower on average compared to graduate student readers ($F(1, 28) = 15.91, p < .001$). Also, the first reading of the story, sentence by sentence, produced longer average reading times compared to re-reading the story sentence by sentence ($F(1, 28) = 7.07, p < .05$). The apparent interaction effect between reader fluency and number of readings was not statistically reliable however ($F = 2.71$).

An analysis of the average standard deviations of reading times also included the factors *words* vs. *sentences*, *graduates* vs. *undergraduates*, and *first reading* vs. *repeated reading*. We observed a statistically reliable interaction effect of text unit with repeated reading ($F(1, 56) = 12.42, p < .001$). The dispersion of reading times narrowed with repeated readings for both reader groups in the sentence-unit condition ($F(1, 28) = 16.20, p < .001$), although the effect was much less pronounced, in fact narrowed to a marginal trend, for word-unit reading times ($F(1, 28) = 3.17, p = .086$, no other interaction effects were apparent, all $F < 2.35$). Other reliable

main effects were observed for text unit ($F(1, 56) = 171.18, p < .001$), repeated reading ($F(1, 56) = 19.06, p < .001$), and a marginal main effect of reader group ($F(1, 56) = 3.83, p = .055$). The primary patterns for interpretation are that dispersion of reading times narrows with repeated readings of the text, and that there is a tendency for graduate readers' reading performance to be less dispersed compared to that of undergraduate readers.

The picture that the conventional analyses present for adult readers is relatively straightforward. Average performance in the word unit condition appears to be at ceiling, being close to what might be expected in a simple reaction time task. In that light, the average reading times to word units are relatively unaffected by fluency or repeated readings, suggesting that all the words presented are represented to the same extent in the mental lexicons of all the readers. Thus differences in reading fluency are not due to whether words are represented in the mental lexicon. Perhaps, though, word reading is automatic for almost all words and most of the readers, so performance stays relatively close to ceiling in the word-unit condition. The sentence unit condition opens up the effect of repeated reading, however, and both graduate and undergraduate readers show a reliable decrease in dispersion of reading times as well as faster reading times overall in re-reading the story.

Apparently, the differences based on reader fluency only emerge when the text is presented sentence by sentence. Consequently we might conclude that the locus of reading fluency is in higher order sentence-level knowledge of linguistic structure. Perhaps the knowledge of how lexical items are used in sentences discriminates between graduate and undergraduate readers, and episodic memory of reading a story is sufficient to either instate, or compensate for, this higher order knowledge. The consequence is seen in faster and more narrowly dispersed reading times.

Fractal Analysis

Now we explore what can be learned about reading performance from the nonlinear structure of the data. One difficulty inherent in the previous picture is that it limits us to inferring differences in reading fluency from quantitative differences in performance, faster vs. slower or wider vs. narrower dispersion. Difficulty stems from underspecification by the data of the theory that we infer from the data. All we can know is that a quantitative difference exists between fluency groups when the reading units are sentences, but the data cannot tell us about any qualitative differences that are entailed. These must be abducted and then tested using additional manipulations to rule out competing hypotheses, effectively playing a form of 20 questions with nature (cf. Newell, 1973).

Nonlinear methods do distinguish among qualitative differences in performance and, apart from the focus on time evolution, these methods make very few assumptions about the system under study, which is an often unappreciated advantage of nonlinear methods. For instance, we need not assume that a stable mean exists, nor that the variance of the data is homogeneous, and no assumptions are necessary about the component processes of which a system is composed nor how those processes interact. The first goal, realized up front, is to characterize the form of the interaction among a system's components, whatever their nature (e.g., see Holden, Van Orden, & Turvey, 2009). This means that the way in which we imagine cognition to be, before we design an experiment, matters much less, initially, to the analysis of reading performance. Reading time data can reveal the nature of the interactions among components, for example, and thus tell us how we should proceed past that point (compare Carello & Moreno, 2005).

The method we illustrate requires several hundred data points at least, collected continuously, for reliable quantification of nonlinear structure (see Holden, 2005, for recommendations). Fractal analyses provide two kinds of information to a scientist about the qualities of the data. On the one hand, the results of a fractal analysis tell us whether a conventional analysis can be justified. A fractal analysis estimates the fractal dimension of a data set, and fractal dimensions that differ reliably from $FD = 1.5$ (the fractal dimension of random white noise) may imply that no stable variance exists and sometimes that no stable average value exists. Conventional analyses grounded in the General Linear Model assume stable mean and variance and that data values are statistically independent, one from another, predicting data series that vary in a pattern equivalent to random white noise.

The fractal dimension of white noise is a good example from which to get a feel for the concept of fractal dimension more generally. The probability that one data point will be succeeded by a value greater than or less than the central tendency of the data is 50–50, we may as well toss a coin. Thus, about half the time a graph will include data points greater than the central tendency, and the other half of the time the data points will be less than the central tendency — so loosely speaking, the data graph fills up about half the space above the central tendency and half the space below the central tendency.

The fractal dimension of white noise, $FD = 1.5$, captures this idea nicely if we imagine that white noise leaks halfway into the 2nd dimension of Euclidean space. The data graph of white noise is not simply a line, random noise is too irregular to be simply a line, but neither is it a 2-D object, it is “halfway” between. Fractal dimensions exist in between the canonical dimensions of Euclidean geometry. Euclidean dimensions are sufficient to capture the highly regular structures of Euclidean geometry, including 0-D points, 1-D lines, 2-D planes, and 3-D spheres or cubes. Fractal objects live between these canonical objects and the data series

presented in Figure 3 is such an irregular fractal object, with $FD = 1.23$. The data graph is not a 1-D line, being highly irregular and aperiodic, and thus leaks somewhat from the 1st dimension into the 2nd dimension, in a manner of speaking.

If a timeseries of data has a fractal dimension reliably different from that of random white noise then a conventional analysis may be unjustifiable and could thus be misleading. The fractal analysis in this case is a test of whether the necessary assumptions of conventional analyses are met. If not then either the variance of the data series is nonstationary or both the mean and the variance are nonstationary, and nonstationary descriptive statistics cannot be trusted to be stable estimates of population parameters (Brown and Leibovich, 2010). The fractal dimension of the timeseries of word reading times in Figure 3, $FD = 1.23$, is very close to a fractal pattern called pink noise (or $1/f$ noise). Idealized pink noise has a fractal dimension of $FD = 1.2$, and thus leaks less into the 2nd dimension compared to white noise. Pink noise indicates nonstationary variance that will grow as more data are collected (e.g., see Van Orden, Kloos & Wallot, 2011).

In addition to the two kinds of information supplied by fractal analysis, there are presently two kinds of fractal analyses applied productively to human event time data. Monofractal analyses estimate a single fractal dimension across an entire data set. They assume that a data series has a stationary fractal dimension. This assumption is not always true however. Human response time data are sometimes monofractal but more often the fractal dimension is nonstationary, changing reliably during the course of data collection.

Multifractal analyses test whether a data set has a stationary fractal dimension or multiple fractal dimensions, varying across a multifractal spectrum. Multifractal analysis is an extension of monofractal analysis. It assesses specifically whether the variation in fractal dimension across a data series is substantially different from what would be expected of monofractal data. Thus a multifractal analysis tests whether variation changes in quality within the same data set. A positive outcome implies that a data series entails multiple different organizations of the system that performs a task.

Holden (2005) is a tutorial introduction to spectral and standardized dispersion analysis to estimate a single fractal dimension of a data set. Peng, Havling, Stanley and Goldberger (1995) describe another monofractal method called detrended fluctuation analysis applied to heartbeat data. Ihlen and Vereijken (2010) discuss multifractal analyses of human data series, including time estimation data and word naming data. Seuront (2010) is a good textbook on multifractal analysis although its examples come from ecology and aquatic science.

Monofractal analysis

Figure 3 illustrates a spectral analysis of one participant's repeatedly measured word-by-word reading times. As noted already, we can expect to satisfy two goals

Figure 3. Top: The time series of one participant's word reading times. Middle: The same time series, but with extreme outliers removed and normalized is subjected to spectral power density analysis (bottom) where sine and cosine waves are fitted to the data to estimate the power at different frequencies. The Power spectrum is plotted on log-log axis to investigate regions of scaling in the time series.

with the results of a monofractal analysis. One result decides the extent to which we were justified in drawing conclusions from the previous conventional contrast of averaged word reading times. The second result concerns whether fractal statistics can reliably distinguish graduate student from undergraduate readers or a first reading of the story from a subsequent reading.

Data preparation. For the adults reading data we followed the guidelines of Holden (2005) in preparing the reading times for the spectral analysis. Although outlier reading times are in principle legitimate data values in nonlinear analyses, sufficiently extreme outliers can bias a spectral analysis, possibly prompting a false rejection of conventional analysis. Thus, extreme outliers were removed to ensure a conservative conclusion in that regard. Before subjecting the data to analysis, word reading times less than 100 ms or greater than 2,500 ms were eliminated, as well as sentence reading times less than 100 ms or greater than 20,000 ms.

Detrending also reduces the likelihood of a false rejection of conventional analysis, and if the pattern of reading times is truly fractal then the removal of outliers and trends should not greatly impact the results (Caccia, Percival, Cannon, Raymond, & Bassingthwaighite, 1997). A safe bet then is to conduct the analysis with and without outliers and trends and trust similar outcomes. In the second step, all the reading times that fell outside of three standard deviations from a participant's average reading time were removed, as well as linear and quadratic trends.

All available techniques to estimate fractal dimension are workarounds of linear methods of analysis and are thus prone to artifacts. To minimize artifacts we use several fractal methods together, checking one result against the others. Standardized dispersion analysis should not be employed when scaling exponents lie between pink and brown noise (that is, between fractal dimensions of 1.2 and 1.1, respectively, see Holden, 2005). On the other hand, detrended fluctuation analysis does often yield better estimates of scaling exponents when the data are not trimmed, because it takes local trends into account (Peng et al., 1995). Also the same trimming and detrending should be applied across the different methods of fractal analysis to produce comparable results (even though detrended fluctuation analysis does not require removal of trends).

Holden (2005) recommends trimming from the beginning or end of the data series down to an integer power of two (e.g., 1024, 2048, 4096, ...) to make the analysis more efficient computationally and to insure that all data sets are of similar length. We have skipped that recommendation because the computational power of available platforms eliminates most concerns about overly long running times and because the present reading time data series have visibly different dynamics initially (e.g., strong initial transient phases in reading times), which we wished to include in the analysis. Finally, the trimmed and detrended data were subjected to spectral analysis, using the Fast Fourier Transformation.

The spectral analysis of each participant's reading times yielded a log-log plot like that portrayed in Figure 3. The Y-axis in the plot tracks the magnitudes of variation as changes in reading times and the X-axis tracks how often changes of particular magnitudes occur. The relation between size of change and frequency of change is estimated by a least-square regression analysis, which quantifies the relation between size and frequency in the slope of the regression line. The end result estimates a scaling relation between size $S(f)$ and frequency (f): $S(f) = 1/f^\alpha$.

The slope estimated in Figure 3 is negative, $-\alpha = -0.76$, which in turn gives us a scaling exponent $\alpha = .76$. This value of the scaling exponent, $\alpha = .76$, lies between the scaling exponents that correspond to idealized white noise (scaling exponent of $\alpha = 0$, slope = 0, fractal dimension = 1.5) and the idealized fractal pattern that we referred to as pink noise (scaling exponent of $\alpha = 1$, slope = -1, fractal dimension = 1.2). The scaling exponents from all conditions can be subjected to conventional inferential statistical analysis (a two-way ANOVA in the present case) to compare average scaling relations from different conditions.

The absolute scaling exponents are of interest to determine whether conventional analyses can be justified. In the example presented in Figure 3, the scaling exponent is within the range of scaling exponents between white noise and pink noise. This range is associated with fractional Gaussian noise, in which variance will increase with increases in sample size (see Van Orden et al., 2011; Wallot & Van Orden, 2011a, for examples). Fractional Gaussian noise has a stable mean that can be informative depending on the hypotheses that are tested.

Otherwise, like other descriptive statistics, scaling exponents (and fractal dimensions) are most informative in contrasts that compare different outcomes, and laboratory performances can yield a wide range of scaling exponent values (Van Orden et al., 2011). Also, just as for other performance measures, the observed variation is due to tradeoffs among task demands, the integrity of the system, participant skills, experimental manipulations, and participants' strategies (Kloos & Van Orden, 2010; Van Orden, 2010).

Results and discussion of monofractal analysis. The analysis of variance yielded no statistically reliable interaction effects (all F values < 1.7, all p values > .20) across the factors *words* vs. *sentences*, *graduates* vs. *undergraduates*, and *first reading* vs. *repeated reading*. The average scaling relations of reading times are portrayed in Figure 4. The most pronounced difference is between word-unit and sentence-unit presentation conditions ($F(1, 56) = 78.98$, $p < .001$). In general, changes in α , in the direction of the $\alpha = 0$ of white noise, have been associated with an increase in sources of involuntary control of task performance (or a reduction in voluntary control, Van Orden et al., 2011, is a review). Keep in mind, though, that the α values close to zero can yet reliably differ from zero, indicating the presence of fractal

behavior and not simply random variation (Gilden, 2001, 2009; Holden, Choi, Amazeen, & Van Orden, 2011).

To test whether the average scaling exponent α of one condition is reliably different from zero, a so-called surrogate data test can be performed. To conduct a surrogate test, each data set is paired with a shuffled copy of itself. The shuffled copy is obtained by randomizing the data points of the original timeseries. This new, randomized timeseries, the surrogate, retains the same central tendency and distribution of the original, but all temporal information is destroyed by the process of randomization. Fractal analysis can now be performed on the surrogate data, and a paired sample *t*-test can be applied to test the scaling exponent values of the original timeseries against the scaling exponent values of their surrogates for each condition. If the *t*-test indicates a reliable difference between the surrogates and the original timeseries, then it can be assumed that the tested timeseries data do not merely consist of random fluctuations, but contain long-range dependencies (see also Holden, 2005).

For example, if we test each sub-condition of word and sentence reading times against their surrogates, then the results indicate that the scaling exponents of the original timeseries are reliably different from the scaling exponents of their associated surrogates.

The observed difference in the values of scaling exponents between the word unit (mean $\alpha = .55$) and the sentence unit (mean $\alpha = .19$) presentation conditions in the adults' reading data can be interpreted as a difference in task demands (see Figure 4). This difference due to different task demands suggests that sentence unit presentation conditions increase the influence of unsystematic task demands on performance measures. If the story presented had had a rhythmic structure, like some poetry, then the change from sentence to sentence could have presented systematic changes that we might have picked up as systematic changes in reading

Figure 4. Average α values for word and sentence reading times bracketed by their standard errors.

times. Otherwise, effectively, unsystematic changes in sentence structure and content are unsystematic perturbations to the measured values that are the input to the spectral analyses, and unsystematic perturbations change the estimates of α values in the direction of random white noise.

Task demands usually vary unsystematically from one measurement trial to the next unless the experimenter designs things otherwise. Measured scaling exponents close in value to the scaling exponent of white noise or values that change in the direction of white noise reflect the increased presence of unsystematic changes in task demands, as the ordered series of measured reading times varies more like a series of randomly selected reading times. Unsystematic differences among sentences are the likely source of the more random pattern, assuming that the sentences that compose the story vary in length, complexity, and linguistic properties, and vary unsystematically in the order in which the sentences appear in the story.

But why would sentences vary more randomly in this regard than words? The syntactic, semantic, and surface properties of the words that compose the story's sentences also vary unsystematically with respect to word order by itself. This is not to say that word order does not matter in sentences, only that word order is unlikely to be more systematically related to reading times than is sentence order. In fact fractal analyses of story-ordered word properties yielded α values close to white noise, namely $\alpha = -0.03$, suggesting that word-order contributes unsystematic variation. These results render an explanation in terms of the difference in task demands alone insufficient.

If the difference between word and sentence conditions is not exclusively in the linguistic or psychological properties of text units, then it may be in the strategy that the reader brings to the task — it may be that the word unit condition also encourages different voluntary control of task performance (rather than a change in involuntary control exclusively). For instance, words are read and understood almost immediately, and more quickly than the time it takes to initiate and follow through with pressing the space bar that is used to reveal new words. Skilled silent reading by undergraduates reaches average speeds of around 300 words per minute (cf., Rayner & Pollatsek, 1994) — close to simple reaction times, and faster than typical performance in a tapping task in which undergraduates produce repeated key presses at their self-selected comfort pace (McAuley, Jones, Holub, Johnston, & Miller, 2006).

An efficient strategy in the word unit condition might be to find a fast pace for repeatedly pressing the space bar and simply read the words apace as they appear across the screen. This strategy explains the data nicely because the average α value in the word unit condition is within the range of values found in tapping performance (Chen, Ding & Kelso, 1997; Delignières, Torre, Lemoine, 2009). The

hypothesis is also consistent with the statistically reliable changes due to number of readings of the story ($F(1, 56) = 5.57, p < .05$) together with the absence of reliable differences between graduate and undergraduate readers ($F < 1$).

The latter result makes sense because graduate and undergraduate readers would not obviously differ in a tapping strategy, and the former result makes sense if the strategy is more reliably engaged in repeated readings when both graduate and undergraduate readers are familiar with the story. For instance it is possible that less familiar words or complicated passages would increase word reading times intermittently, but sufficiently, in the first reading to supply unsystematic perturbations of measured values. A subsequent re-reading of the story could dampen these perturbations as repetition of story elements yields consequently faster comprehension of the now familiar words and sentences.

The previous ideas gain credibility in a closer look at adult reading performance, in the separate analyses of *word by word* vs. *sentence by sentence* presentations of the story. The estimated α values across word-by-word reading times are portrayed in the leftmost bar graph of Figure 4. An analysis confined to the word unit condition reveals a reliable change in average α values with repeated readings of the story ($F(1, 28) = 4.79, p < .05$). The average α values change from values closer to white noise ($\alpha = 0$) toward values closer to pink noise ($\alpha = 1$), and the same pattern of change occurs reliably for both graduate and undergraduate readers, which do not differ statistically otherwise, and there is no statistically reliable interaction between these factors (both F s < 1.22). These changes are unique to the word unit condition because there are no reliable differences among the much smaller α values of sentence reading times (all $F < 1.14$). It makes sense then that word unit presentation encompasses a tapping strategy in self-paced reading.

Multifractal analysis

Monofractal analyses assume that the pattern of variability across a time-ordered data set is sufficiently characterized by a single stationary α value. This is certainly not always the case, and perhaps it is not even often the case. The more general outcome appears to be nonstationary α values that fluctuate within the course of a task performance (Ihlen & Vereijken, 2010). Multifractal analysis addresses this question of the quantity and the heterogeneity of α values that describe a data series (Brown & Liebovitch, 2010).

Mathematically, the tools of multifractal analyses are formal extensions of the tools of monofractal analyses. The monofractal α value can be vaguely thought of as the central tendency of the multifractal case, and the spectrum of α values that result from a multifractal analysis can be seen as a dispersion of α values around

that central tendency — keeping in mind that α values differ in the implied quality of dynamics, not merely quantity. A key statistic of interest is the width of the multifractal spectrum (as in a loose analogy with a standard deviation or standard error of a mean). If the width is sufficiently small, the monofractal α value is a good enough estimate of the fractal properties of a data set. If the multifractal spectrum is sufficiently wide then it would be a mistake to ignore this essential character of the data, that its fractal properties are changing during the task performance and indicate that the underlying dynamics are changing in quality (Riley & Turvey, 2002).

Multifractal phenomena are unabashedly complex in the strict formal sense of complexity science, implying interaction-dominant dynamics among the component processes of the reading performance (Ihlen & Vereijken, 2010). The presence of multifractal dynamics in data may also imply that a studied system self-organizes its behavior to suite the requirements of the task. Interaction-dominant dynamics allow the component processes of a system to change each others dynamics perpetually as they interact (Van Orden, Holden, & Turvey, 2003). This creates interdependence among component processes, which allows context to constrain changes in all interacting processes, becoming constitutive of the self-organization (Van Orden, Kello, & Holden, 2010).

Results and Discussion of Multifractal Analysis. We conducted multifractal analyses of the reading time data, applying two prominent analysis techniques. Multifractal Detrended Fluctuation Analysis (Kantelhardt, Zschiegner, Koscielny-Bunde, Havlin, Bunde, & Stanley, 2002) and Multifractal Continuous Wavelet Transformation (Ihlen & Vereijken, 2010) were used together, allowing us to compare outcomes and enhance reliability by ensuring common outcomes. We report the average widths of the multifractal spectra estimated using the multifractal continuous wavelet transformation.

Figure 5. Average widths of multifractal spectra computed from stimulus-ordered series of word and sentence reading times, bracketed by the standard errors of the means.

The single reliable interaction was between *word vs. sentence* unit presentations and *single vs. repeated reading* of the story ($F(1, 56) = 9.69, p < .05$; all other interaction terms yielded $F_s < 1.29, p > .25$). Figure 5 portrays the average widths of the multifractal spectra for word and sentence reading times of the adult's data. The average width of the multifractal spectrum from word-unit reading times shrinks reliably with repeated readings for both reader groups ($F(1, 28) = 8.92, p < .01$). There was no effect of repeated readings or reader group on the multifractal spectrum of sentence unit reading times (all $F < 1.53$).

Fractal and multifractal characteristics of sentence reading times are not reliably different for different reader groups or repeated readings. Monofractal exponents are close to white noise in the sentence-unit condition and the widths of the multifractal spectra are narrow. Thus sentence reading times are fairly well characterized by their monofractal exponent, which suggests the presence of unsystematic perturbations to reading times due to unsystematic variation of sentence properties, in the story order of sentences. Applying the surrogate test to the multifractal widths of the sentence data (i.e., we randomize the timeseries, conduct multifractal analysis on these surrogates, and compare the outcomes with multifractal results obtained from the original timeseries), the spectra widths observed in sentence reading are not reliably different from zero — none of the sentence conditions show reliably multifractal behavior.

However, both the monofractal and multifractal analyses are sensitive to the conditions of word reading times. Monofractal exponents increased with repeated readings and the width of the multifractal spectrum decreases with repeated readings. Both these outcomes make sense with respect to the hypothesis that the word unit condition allows a special tapping strategy, to tap quickly and read words apace. Again, graduate and undergraduate readers would not obviously differ in a tapping strategy, and the strategy is more reliably engaged in repeated readings when both graduate and undergraduate readers are familiar with the story. When the story is familiar, the influence of story idiosyncrasies is reduced allowing a more coherent tapping strategy and a fractal pattern closer to pink noise, deviating to a lesser degree from pink noise.

Recurrence quantification analysis

Recurrence quantification analysis begins with phase-space reconstruction. The phase space portrait is inferred from the 1-D data series of reading times. This phase space is usually higher in dimension than the 1-D data series. The ordered reading times are plotted against themselves by forming one or more data series from the original data by creating ordered pairs (triplets, quads, ...) of data points

with a point that lays it in the ordered data series by a chosen constant value — a delay parameter. This constructs a higher dimensional portrait of the system dynamics. The constructed space is used, in turn, to define structure in a 2-D recurrence matrix in which only equivalent reading times at equivalent locations in the phase space are marked. Webber and Zbilut (2005) is an accessible introduction to phase-space reconstruction in RQA; the mathematical rationale and proofs can be found in Takens (1981).

The recurrence matrix is formed by following the order in which the data points were collected on both axes of a matrix, and then marking each point in the matrix at which equal reading times recur at the same location, along a shared trajectory, in the higher-dimension phase space. The recurrent reading times are thus located in the neighborhood of equivalent values along a shared trajectory of the space that was formed using lagged reading times.

Recurrence quantification analysis (hereafter RQA) quantifies the geometry of the dynamics represented in the recurrence matrix. RQA yields variables that quantify the dynamics of a hypothetical system that could have produced the reading time data. The central concept of recurrence concerns how often a system repeats its behavior. Data that follow a regular pattern of variation repeat the same behavior often, producing a lot of recurrences and a simple regular geometry in the recurrence matrix. In more complex data series, recurrences become less frequent and the pattern of recurrences itself can change across the recurrence matrix. Recurrence quantification analysis quantifies these aspects of the matrix geometry producing variables that can distinguish among different data sets.

The number of dimensions of the constructed space also estimates the number of active system components (i.e., active degrees of freedom) producing the variation in reading times. RQA works because continuously updating feedback creates coupling among the active degrees of freedom, sufficient to create interdependence among the components. Interdependence means that each change in each component is reflected simultaneously among all the components. Interdependence thus imbues the changes in each component with information about how the other components are changing.

Because everything is connected to everything else during behavior we can use the information latent in 1-D data to recover the higher-dimensional dynamics of the hypothetical dynamical system in question. Marwan (2011b) discusses contemporary issues in applying the analysis, including how many data points are necessary to conduct Recurrence quantification analysis reliably. The range is wide: sometimes, several hundred data points suffice, sometimes tens of thousands of data points are needed, depending on the complexity of the data series and the dynamics of the performance of interest. As a general rule, if the results

of a recurrence analysis are based on less than 400 data points, they should be interpreted with caution and appropriate skepticism.

Data Preparation. In a report of RQA analyses, it is import to describe how the data were prepared before the analysis, how many data points each data series contained, the parameter values that were picked, and which parameters were held constant across the analysis. The data should be normalized to begin with, setting the parameters *norm* and *rescaling*, usually to the Euclidean norm and the maximum distance rescaling option. Apart from special circumstances, the most important thing about norm and rescaling is to keep them constant across the analyzed data sets that are compared.

Also, RQA mandates the presence of each reading time, kept in the same order as the reading times were collected. If any data points are eliminated then their adjacent reading times, on either side, are moved closer together. This distorts the phase-space portrait, which distorts the recurrence plot that, in turn, distorts the quantities computed from the recurrence plot. Hence, only non-trimmed and non-detrended data series are subjected to RQA — with one exception:

Suppose that a recorded data series contains data from two different contiguous events, for example eye-position measures recorded by an eye-tracker during calibration versus eye-position measures taken right after, during a reading experiment. In such a case, the calibration part of the data series can be removed up to the point where the eye-movements of reading begin. If the data series is incomplete (that is, it is truly missing data points), it is advisable to estimate the distribution of the data series and randomly draw values from the distribution to fill in the gaps. It would not be advisable to insert systematic values for missing data points, such as the averaged values of adjacent data points. Using averaged values risks inserting spurious structure into the recurrence plot.

Figure 6 is a cartoon illustration of a 1-D data series converted into a 3-D phase-space portrait. The data plot in the upper-left quadrant of the figure has time (presentation order) on the X-axis and the data measurement on the Y-axis. The data plot is also marked with three dots to indicate the lag delay for re-graphing the data in a 3-D portrait. The delay parameter is the distance, on the X-axis, between the lagged points. The lagged points are used to re-plot the data series (against itself) in the higher-dimension phase space.

The delay parameter is selected so as to minimize the interaction between points of the measured timeseries (maximize the information available across the lagged points). This parameter is commonly estimated using the first minimum of the autocorrelation function or else using the mutual average information function of the data series. With the delay parameter in the example set at four, the data series can be replotted. The Y-value of the left-most dot in Figure 6 defines the value on a new X-axis in 3-D, the Y-value of the middle point is the value on a new

Figure 6. From a one dimensional time series to a three-dimensional reconstructed phase-space to a recurrence plot. The top row illustrates how a time series plotted against itself at time t_4 , t_8 , and t_{12} yields coordinates for a three-dimensional phase-space portrait of that time series (i.e., $x = t_4$, $y = t_8$, $z = t_{12}$). The bottom row illustrates how a neighborhood (indicated by the red circle) in the phase-space gives rise to one or more points in recurrence plot.

Y-axis, and the Y-value of right-most point is the value on a new Z-axis, resulting in the (x,y,z) point in 3-space pictured in the upper-right quadrant of the figure.

The dimension parameter defines the number of dimensions of the phase-space portrait that is (re-)constructed (the number of times the timeseries is plotted against itself with the specified delay). The first minimum of a nearest-neighbor function is usually used to estimate the number of dimensions needed. The data points that are used in this procedure (delay $\times$ dimension) are lost for the quantifications. Hence if dimension is a high value and/or delay is a large value, then the remaining data points have to be sufficient in number to reliably estimate quantities from the recurrence plot.

With all the data replotted in 3-space (the spiral shape of the upper-right quadrant), a neighborhood criterion is chosen — the radius parameter — portrayed in the lower-left quadrant by a circle. The radius parameter sets a threshold within which points in the reconstructed phase-space portrait are counted as equivalent recurrent points. The radius is chosen conservatively to define the equivalence neighborhood around each trajectory in phase-space, which in turn decides which of the nearby points should be counted as intersecting trajectories (given a little noise). Generally, the radius parameter is chosen so that the recurrence rates do not exceed 5%.

Usually, the radius size is kept constant for all the data sets that are compared. However, it is sometimes more useful to keep recurrence rate constant, letting the radius vary. This is particularly useful if no radius can be found that brings all the data sets into the range of 1–5% recurrence rate. In Figure 6, the three, vertically stacked, data points, located one above the other, within the radius of the middle point, are deemed equivalent (with the middle value)— they are recurrent points on intersecting trajectories. Consequently, when we plot the recurrence matrix a dot can be inserted to mark all possible pairs as recurrent points.

Five recurrent points are illustrated in the recurrence matrix in the bottom-right quadrant of Figure 6. The recurrence plot is fully fleshed out by marking each intersecting point of recurrent values with a dot. With all these dots in place the larger structures of the recurrence plot become visible as lines of dots, 'empty spaces,' and larger geometrical structures. The visible structure provides the basis with which to quantify the nonlinear variables of the recurrence quantification analysis.

For the reading time data, the entire data series of 13,925 data points for words and 1,042 data points for sentences (adults), and 1,099 data points for words (children) was subjected to RQA without trimming or detrending. The data were transformed into z-scores prior to the analysis. The parameters were estimated for each data set individually. Mutual-average information was used to estimate delay parameters ranging from one to 54, and false-nearest-neighbor analysis estimated embedding dimensions ranging from three to 11.

We used the same delay value = 2 for all data sets, dimension = 8 for word-unit readings and dimension = 4 for sentence-unit readings. For the adult reading data, the radius parameter was kept constant across all data sets within each text condition. That is, different radii were used for word and sentence reading times, but they were kept constant within each of the two conditions.

RQA yields multiple quantitative outcome variables, and more are currently under development. Not all outcome variables are always applicable or meaningful for every data set (see Marwan, 2011b). Hence, one must pick measures that make sense in the context of the data at hand and the research questions to be answered. Some measured values such as recurrence rate (%RECCurrence) or the amount of

deterministic structure in the data (%DETerminism) are meaningful in most contexts, and for this illustration using the reading data we chose %REC, the size of the radius parameter (RAD), and %DET, to illuminate self-paced reading.

Results and discussion of recurrence quantification analysis. All of the analyses we present in the following sections were conducted using Norbert Marwan's 'Commandline Recurrence Plots' software (Marwan, 2011a). Figure 7 portrays %RECCurrence and %DETerminism results for word and sentence units of the adult's reading data. No reliable main effects or interactions were observed for repeated readings or reader group (all $F < 1.33$) for word response times. However, the sentence unit data yielded a marginally reliable interaction effect between repeated readings and reader groups for %DET ($F(1, 108) = 2.93, p = .098$), and the same pattern is replicated in %REC. Recurrence variables each provide independent information about a data series, so similar patterns are replications, yet marginal effects. The present pattern in %REC and %DET is a decrease for graduate readers in a repeated reading, while the same factors increase for undergraduate readers in a repeated reading.

Higher %RECCurrence (for the same value of the radius parameter) indicates that the behavior is less spread out in the phase space, perhaps less noisy. %DET equals the number of diagonally adjacent points in the recurrence plot divided by the total number of recurrent points, which estimates the degree of order in the data, the extent to which readers fall into coherent trajectories of similar reading times across text units.

The contrast between results from word and sentence unit conditions corroborates the hypothesis that word by word reading is of a different quality than sentence by sentence reading. The word unit condition resulted in similar outcomes for undergraduate and Ph.D. candidate readers, consistent with the hypothesis that they find a common strategy of quick taps of the spacebar while reading words apace. The cross-recurrence quantification analysis will provide another, stronger test of this hypothesis.

The sentence-unit condition can now be seen to be the more interesting condition because it distinguishes between Ph.D. candidate readers and undergraduate readers (re-reading a text has apparent opposite effects on the two categories of readers across the RQA variables). The opposite directions of the effects may imply different qualities of effects and different qualities of readers — fluent undergraduate readers versus über-fluent Ph.D. student readers.

The success of applying recurrence quantification analysis to reading times of adults and children to investigate systematic differences between repeated reading, reader groups, or grades has so far been limited. However, there is another way to test how the influence of the different reading conditions impact the performance of readers using cross recurrence analysis. Cross recurrence analysis can be used

Figure 7. %REcurrence, %DETerminism and Average TrappingTime, with standard error bars, for word and sentence presentations conditions.

to look at commonalities among readers in a group that are due to the common accommodation of task demands, including how the text-unit conditions affect readers similarly or differently.

Cross-recurrence quantification analysis

Cross-recurrence analysis extends recurrence analysis in much the same way that a correlation between two random variables extends auto-correlation of a single random variable. In the latter case, we evaluate a data series against itself, and in the former case against another data set. In this light, cross-recurrence analysis (hereafter CRQ) can be thought of as a kind of nonlinear correlation analysis developed to test whether, and the extent to which, two systems produce dynamics in common (Shockley, Butwill, Zbilut, & Webber, 2002). If they do, then common sources of constraints are implicated, where a change in constraints is associated with a change in the degrees of freedom of trajectories in the state space. More tightly constrained dynamics have fewer degrees of freedom and less tightly constrained dynamics have more degrees of freedom.

The procedure for conducting the CRQ analysis is nearly the same as that for conducting RQA, the primary difference being that two data sets are contrasted instead of only one. That being the case, it can be crucial to normalize both data sets prior to analysis (for example by computing z-scores) so that they share a common scale. Shockley (2005) gives a good introduction to cross-recurrence quantification analysis (and see also Marwan, 2011b). We used CRQ to compare the shared dynamical structure among readers who read the same text under the same conditions. So, for instance, we asked whether two undergraduates produced shared patterns of reading times because they read the same text, advancing it identically, unit by unit in spacebar presses. The idea behind this analysis is that readers may possibly entrain to a text in self-paced reading or adopt the same strategy for advancing the text in the self-paced reading task. Such commonalities may yield parallel dynamics that can be captured in CRQ.

For the adult reading times, each condition included eight readers and we examined all possible pairings, which yielded $8!/(2! \cdot (8-2)!) = 28$ possible pairings in each cell of identical reading conditions. The data were transformed into z-scores prior to the analysis. The parameter settings for the CRQ were identical to those chosen for the RQA (delay value = 2 and dimension = 8 for all data sets, including both word and sentence data, radius parameters were set to yield averages of 2% recurrence for each text unit condition).

Results and discussion of cross-recurrence analysis of adults' word reading times. As can be seen in Figure 8, %REC and %DET drop reliably for both reader groups with repeated reading of word units: %REcurrence ($F(1, 108) = 13.61, p < .001$), %DETerminism ($F(1, 108) = 24.84, p < .001$). These precipitous drops in the CRQ measures all pertain to the constraints available that might support shared dynamics among readers. Apparently, sources of constraint available in the first reading of the text, in the word unit condition, have a reduced effect in the subsequent reading.

Figure 8. Cross recurrence analysis of word and sentence reading times with standard error brackets.

The decreases in shared %REC and shared %DET indicate that less shared and more idiosyncratic dynamics emerged with a subsequent word-by-word reading. Re-reading a story word-by-word could disperse trajectories in a reader's reconstructed phase-space and reduce the likelihood that readers share trajectories of reading times. This would decrease %REC and %DET in the CRQ variables of the re-reading. In other words, one source of the increasingly idiosyncratic performances could stem from data that are more randomly dispersed due to extraneous sources of random noise and this possibility fits the outcomes of the CRQ. But it is not supported by other data.

For instance, the standard deviations of reading times are apparently smaller for repeated readings, word-by-word, suggesting a more stable performance and a reduction of extraneous noise. Also the monofractal analyses discovered a less random pattern of noise in the repeated reading, together with a narrower spectrum of multifractal dynamics. And the RQA revealed trends toward more determinism in re-reading. These results together rule out extraneous sources of random noise as the reason for less shared structure in reading times across participants (cf. Riley & Turvey, 2002).

Possibly, these results imply a more flexible coupling between readers and the self-paced reading task with repeated reading (cf. Van Orden et al., 2011). We may speculate that participants eventually discover the tapping strategy in all word-by-word presentation conditions. This strategy hypothesis is most strongly supported by the high rates of %DET in both individual RQAs and joint CRQ analyses, and the fact that similar results were found for both English literature, Ph.D. students and undergraduate readers. In the first reading, however, the text supplies perturbations to this strategy in rare words, or difficult passages, or in other aspects that compose unsystematic perturbations of reading performance. Re-reading the

same text word-by-word reduces these sources of perturbation, which are sources of random noise in individual performance.

These same sources of shared perturbation are also sources of shared dynamical structure in comparisons between individuals, bolstering the CRQ variables in the first reading. A re-reading of the story (word-by-word) will benefit from the reduction of perturbations, as indexed by reduced standard deviations, less white noise structure, a narrower multifractal spectrum, and trends of increasing RQA variables. However, the re-reading also reduces the shared dynamical structure that rare words, difficult passages, and other surprises provide to readers. And it is the reduction of these sources of shared structure that reduce the CRQ variables.

Results and Discussion of Cross-Recurrence Analysis of Adults' Sentence Reading Times. The sentence unit condition yielded reliable 2-way interactions between reader group and number of readings for both CRQ variables: %REcurrence ($F(1, 108) = 11.62, p < .001$), %DETerminism ($F(1, 108) = 6.71, p < .05$). %REC and %DET appear to stay about the same, relatively, for graduate readers, whether reading the story for the first time or re-reading the story. In contrast, %REC and %DET increase reliably when undergraduate readers re-read the same story. This outcome would be expected if graduate students' reading performance is always close to a performance ceiling, even reading a new story, sentence-by-sentence, for the first time. That is, graduate students in English literature are über-fluent, compared to fluent undergraduate readers.

The hypothesis that Ph.D. students in English literature are über-fluent is patently intuitive and is supported across the board by the total pattern of outcomes. Throughout all the analyses of sentence-by-sentence reading, graduate students produced more similar scores on all variables between first and repeated readings, when compared to undergraduates. For both RQA and CRQ variables, graduates

Figure 9. Cross recurrence analysis of adults' sentence reading times with standard error brackets.

student scores never differ reliably in the statistical analyses and the apparent changes are always in the direction opposite to that of undergraduate readers. Undergraduate's RQA and CRQ scores always increase, often reliably, after the first reading. This would be the case if the undergraduates' sentence reading times are more uniform in the repeated reading, which may result from a reduction of text-based perturbations affecting the first reading.

The results of sentence unit reading times make sense if the graduate students are near the performance ceiling of reading fluency, which we can only see clearly when first and subsequent readings are compared and the presentation units are right (sentence units). Wallot & Van Orden (2011b), found a similar outcome using phrase-unit presentations. Reading times to sentence units are rate limiting compared to spacebar pressing, which is not true of reading times to word units, and so sentences allow the more telling portrait of reading performance. Also, the performance of undergraduate readers became more stable and deterministic when rereading the story sentence-by-sentence. They repeated more of the same reading times at the same locations in the phase space of their performance dynamics.

Study 2: Children's reading times

Figure 10 presents average reading times and standard deviations for the children's word reading times. As can be seen, neither average reading time, nor average standard deviation statistically distinguishes between the performance of children in different grades (both $F < 0.15$).

Word reading times and standard deviations do not differ for 2nd, 4th, and 6th graders. The standard error of the statistics seems to be higher for 2nd graders,

Figure 10. Children's average and standard deviation of word reading times in seconds bracketed by standard error.

indicating that they might be a more heterogeneous group of readers, but no further conclusions can be drawn from their performance.

Monofractal analysis

Data Preparation. For the children's word reading times, trimming is a more complicated matter. When a constant trimming criterion is applied, the data of the younger children's performance is much more heavily impacted compared to that of older children (see also Kiefer, Gresham, Wallot, Kloos, Riley, Shockley, & Van Orden, 2011). Furthermore, it seems realistic that younger children will perform less well on the reading task, so that longer response times that might otherwise be seen as outlier performances are actually part of the normal performance pattern. To avoid introducing or leveling differences between the different groups of children through data trimming, the raw data series were used for the analysis.

Results and Discussion of Monofractal Analysis. An analysis of scaling exponents of the children's reading times results in no statistically reliable differences in the reading performance between different grades when spectral analysis (PSD) was used to estimate scaling ($F < 1.85$). However, when detrended fluctuation analysis was used (DFA), we observed a marginally reliable effect of readers

Figure 11. Average scaling values for children's word reading times bracketed by their standard errors. The left panel shows the average alpha exponents obtained from spectral analysis (= power spectral density, PSD). The right panel shows the average Hurst exponents obtained from detrended fluctuation analysis (DFA) of the same data. An alpha value of 0 or a Hurst exponent of 0.5 would both indicate white noise, while an alpha value of 1 or a Hurst exponent of 1 would both indicate pink noise fluctuations in the reading times. The Hurst exponent (H) can be converted into the alpha exponent (α) by $\alpha = (H - 0.5) * 2$, if the exponents lie in the range between white and brown noise (i.e., $0.5 < H < 1.5$, and $0.0 < \alpha < 2.0$).

age ($F(2, 10) = 3.17, p = .086$), indicating that reading times of older children contained weaker long-range dependencies compared to the reading times of 2nd graders. Also, as can be seen in Figure 11, detrended fluctuation analysis yields smaller standard errors compared to spectral analysis, indicating that it can handle the more heterogeneous data set better — given, of course, that the statistical tendency observed will turn out to be a reliable difference as the number of participants is increased.

The scaling exponents of children's reading data give a first hint at the developmental changes in reading performance between 2nd and 6th graders. The fluctuations in 2nd grader's reading times appear to possess a higher degree of long-range dependencies compared to the reading times of 4th and 6th graders, which are essentially white noise.

As with the adult readers, text features may contribute unsystematic variance to a reader's performance. If so, the 2nd graders' reading performance seems somewhat more independent of word properties, exhibiting a greater degree of intrinsic variation, like what we saw in the adults' data due to a simple tapping strategy (see above). However, the overall reading times are much slower than those of adult readers, making it unlikely that a strategy of faster word recognition drives their performance.

It rather seems that the word reading performance of children is much more akin to the sentence reading performance of adults. The reading performance of 4th and 6th graders could then be interpreted as being guided or driven by text features compared to the 2nd graders, which could be a developmental achievement in reading skill. However, the higher scaling exponents of 2nd graders hold a certain degree of ambiguity as well, since they make their performance look more like that of adults than that of 4th and 6th graders in absolute value, while we would suspect that the similarities in reading skill are just the opposite. Further analysis might resolve this ambiguity.

Multifractal analysis

Results and discussion of multifractal analysis. Figure 12 gives an overview of the multifractal spectra widths of the children's reading data. The differences in the multifractal spectra mirror the differences observed in the monofractal detrended fluctuations analysis, but the differences between grades are not reliable ($F < 1.64$).

The children's reading times do possess significant multifractal structure, indicating that the evolution of reading times contains inhomogeneous variance. However, children do not differ across ages in the amount of heterogeneity of the reading times they produce. Multifractal analysis quantifies the deviations from

Figure 12. Average widths of multifractal spectra computed from stimulus-ordered series of the children's word reading times bracketed by standard error of the mean.

homogeneity in a fractal data series. The untrimmed data series of the children tend to produce larger multifractal spectra compared to the adult data. This is due to the very long response times in the children's data sets. However, all of the data sets entered the analysis untrimmed (as raw data), so comparability of the data sets can be assumed. Comparing performances from diverse populations requires that data preparation be equated carefully.

Recurrence quantification analysis

Data Preparation. All of each child's 1,099 word reading times were subjected to RQA without trimming or detrending. The data were transformed into z-scores prior to the analysis. RQA parameters were estimated for each data set individually. Mutual-average information was used to estimate delay parameters ranging from one to 12, and false-nearest-neighbor analysis estimated embedding dimensions ranging from two to 6.

We chose to use the same delay value = 2 for all data sets, and dimension = 4. However, the variability between individual children's data sets was very large, so that we could not find a meaningful radius size that would yield recurrences for every data set. Hence, we chose to fit the radius parameter individually to each child's reading timeseries, such that each timeseries resulted in 2% of recurrent points (as opposed to the adult data, where each condition was assigned a single radius parameter, resulting in an average of 2% recurrence across all data sets of that condition, but allowing differences in the percentage of recurrent points between data sets).

For data sets that are characterized by roughly the same parameters, it makes little difference whether each data set enters the analysis with its individual

parameter settings, or with parameter values chosen to be identical for all the data sets. However, if one decides to fit the radius to each data set, it has to be checked whether the overall differences in other recurrence variables remain similar. Under some circumstances, it can happen that a difference between two conditions results in higher determinism for one condition compared to another condition when the radius parameter is kept constant, but reverses (i.e., lower determinism for the first condition compared to the second condition) when each data set is assigned its own radius parameter, keeping the percentage of recurrent points constant. Under such circumstances, it might be wise to interpret only differences in the radius parameter (which we will describe in more detail in the following section).

Results and Discussion of Recurrence Quantification Analysis. Figure 13 presents the results for the average radius parameter and the average %DEterminism of the children's reading times. Since we fixed the percentage of recurrent points (%REC) for each data set to the same value, namely 2%, a direct comparison of the data sets regarding %REC is not interesting. Instead, we focus on the differences in the radius parameter.

RAD and %REC have a positive relationship, in that %REC rises with an increase of the RAD parameter, because points in phase space that are farther apart are more likely to be counted as lying in the same neighborhood. However, when %REC is equated for each data set to the same value, while there are in fact systematic differences among data sets in %REC, then the differences in the RAD parameter will contain information about these differences: Imagine RAD is kept constant and on data set has higher %REC compared to data another set. If we then choose a RAD for each of the two data sets so that they have the same %REC, we would have to use a bigger RAD for the data set with the lower %REC.

Thus, when %REC is equated across data sets, a higher RAD parameter indicates that a data set would have had lower %REC, if RAD had been kept constant. Hence, the left graph in Figure 11 suggests that the word reading performance of 2nd graders possesses less recurrent structure compared to the performance of 4th and 6th graders, and that the performance of 4th graders possesses more recurrent structure than the performance of 6th graders. However, none of the differences are statistically reliable ($F < 1.01$). Also, none of the average differences observed in %DET are statistically reliable ($F < 2.07$).

Similar to the adults' reading data, RQA did not yield reliable distinctions in the reading performance of children in different grades. Hence, we apply cross-recurrence analysis to investigate commonalities in the evolution of reading times among children within each grade.

Figure 13. RADius size and %DEterminism, with standard error bars, for word reading times of 2nd, 4th, and 6th graders.

Cross-recurrence quantification analysis

Just as with the adult reading times, we investigated all possible pairings of child participants for each grade, resulting in $3!/(2!(8-2)!) = 3$ paired data sets of 2nd graders, and each $5!/(2!(5-2)!) = 10$ paired data sets of 4th and 6th graders. Also as with the adult reading data, the values for delay and dimension were those used in the RQA. The radius was again fitted to each pair of data sets to result in 2% cross-recurrence.

Results and discussion of cross-recurrence analysis of children's word reading times. The CRQ co-evolution of children's word reading times decreased reliably in RADius for the higher grades ($F(2, 20), 3.36, p = .055$) and increased reliably in shared %DET ($F(2, 20), 4.51, p < .05$). That is, the shared dynamics in reading performance among the children become stronger with age as their reading skill increases and as their performance became more closely tied to text properties.

The results corroborate our previous speculation in the context of average reading times and the changes in monofractal structure: That is, the word reading data of young children in a self-paced reading task are akin to that of sentence reading data of adults. Children's word reading times are sufficiently slower to be better captured by the times between button presses.

Changes in CRQ variables across grades 2, 4, and 6 also align well with changes in CRQ variables of undergraduate readers from the first reading to the repeated reading: Repeated reading may increase the undergraduate readers' fluency for the specific text that is re-read (Samuels, 1979), and this gain in skilled reading reveals itself in the tighter coupling of the reader's performance to the structure provided by the text, as indicated by the increases in the CRQ variables.

Figure 14. Cross recurrence analysis of children's word reading times with standard error brackets.

Likewise, children's performances become more tightly coupled to text properties with increased reading skill.

The contrast between the performances of children and adults also suggests that a life-span metric of reading fluency may not require identical testing conditions across the life-span. The development of reading skill may also require complementary changes in the tests used to assess it. At least the comparison of the children's and adults' performance suggests that a change in reading skill can potentially change the very task that is used to assess it.

Conclusions

Both conventional and nonlinear analyses evaluate self-paced reading performance, but they do not reveal the same outcomes either in kind or magnitude. Also, all else equal, the results of the nonlinear analysis must always trump analyses that derive from the General Linear Model, in the sense that the nonlinear results will apply more generally than the more narrowly conceived linear analyses (analyses that assume continuity and proportionality, see Mitchell, 1999).

For example, the monofractal analysis identified the patterns of variation across individual reader's reading times as fractional Gaussian noise, which has indefinite variance. For human event time data this means that the larger the sample of data collected, the more extensively a project will explore ranges of values that are very large (slow) event times, due to a pronounced skew toward slow times.

When data sets do not fit the assumptions of the General Linear Model, many scientists resort to the assertion that ANOVA (for example) is robust in the face of heterogeneous variability — sometimes as a last resort it seems. Nonlinear analysis techniques open up new options to those who investigate complex data sets or

happen to have collected data that do not satisfy the assumptions of the General Linear Model.

Self-paced reading times exhibit pronounced heterogeneous variability (see Figure 3). Consequently, although the data have a defined population mean in principle, it may be impossible to estimate this mean reliably in practice. Indefinite heterogeneous variance implies sample variances that are unstable. The sample variance will impact the empirical estimate of the mean of repeated measures taken from a single participant, like the data from the self-paced reading task. Differences between means remain meaningful, nonetheless, although what they mean is more readily derived within the prior context of the nonlinear analyses. The nonlinear analyses are more general, the linear analyses are the special cases.

Nonlinear methods of analysis may also trump linear methods by revealing differences in kind as well as in magnitude. For instance, the conventional analyses of self-paced reading data could easily have led us astray without the corrective force of the nonlinear outcomes. The conventional analyses suggested to us that über-fluency has a causal basis in — is "localized" in — sentence knowledge, which turned out to be misleading. The conventional analysis had lead us to ask: What is it about sentences versus words that distinguishes über-fluency from ordinary fluency?

In actual fact, the adult participants were doing one kind of task in the word-unit condition and a different kind of reading task in the sentence-unit condition. And it was the nonlinear analyses that clarified how and where we were misled. Thus the better question turned out to be: What strategy distinguishes word-by-word presentations from sentence-by-sentence presentations?

With an answer to that question we could refute the false promise of the conventional results, and the answer was got from nonlinear analyses of reading plus previously published nonlinear analyses of tapping and simple reaction time. How much longer might it have taken to recover from the error, without the help of nonlinear methods? Consider that word-by-word presentation to adult readers remains a standard condition of self-paced reading experiments. Here we found that it gauges the performance of children and college student participants very differently.

For adult readers, it was the sentence-by-sentence reading conditions that distinguished über-fluent Ph.D. students in English literature from fluent undergraduate readers. And the fleshed out nature of the distinction came almost exclusively from gains to be had by fluent undergraduate readers in a repeated reading of the same story. Those gains were not in superficial aspects of performance such as reading speed, however, they were demonstrated in the more subtle aspects that refer to stability, order, and determinism.

Acknowledgements

We thank Anna Haussmann for help with the preparation of the manuscript. Preparation of this chapter was supported by NSF grants to Guy Van Orden (BCS #0642716; BCS #0843133; DHB #0728743).

References

- Accelerated Reader (2011). *Book Finder US*, retrieved August 15, 2011 from <http://www.arbookfind.com/bookdetail.aspx?q=59508&l=EN&slid=194455065>
- Bassingthwaighte, J. B., Liebovitch, L. S., & West, B. J. (1994). *Fractal physiology*. New York: Oxford University Press.
- Brown, C., & Liebovitch, L. (2010). *Fractal analysis*. Los Angeles: Sage.
- Caccia, D. C., Percival, D., Cannon, M. J., Raymond, G., & Bassingthwaighte, J. B. (1997). Analyzing exact fractal timeseries: Evaluating dispersional analysis and rescaled range methods. *Physica A*, 246, 609–632.
- Carello, C., & Moreno, M. (2005). Why nonlinear methods? In M. A. Riley, G. Van Orden (eds.), *Contemporary nonlinear methods for behavioral scientists: A webbook tutorial* (pp. 1–25). Retrieved August 23, 2009, from <http://www.nsf.gov/sbe/bcs/pac/nmbs/nmbs.jep>
- Castillo, R. D., Van Orden, G., & Kloos, H. (2011). The embodiment of time estimation. In A. Avatakis (Ed.), *Multidisciplinary Aspects of Time and Time Perception* (pp. 196–207).
- Chen, Y., Ding, M., & Kelso, J. A. S. (1997). Long term memory processes (1/f^α type) in human coordination. *Physics Review Letters*, 79, 4501–4504.
- DeGrado, L. P. (2003). *The Arelis complex*. Retrieved August 23, 2009, from <http://www.authors-den.com/visit/viewshortstory.asp?id=9632>
- Delignières, D., Torre, K., & Lemoine, L. (2009). Long-range correlation in synchronization and syncopation tapping: A linear phase correction model. *PLoS ONE*, 4, e7822. doi:10.1371/journal.pone.0007822
- Gao, J., Cao, Y., Tung, W.-W., & Hu, J. (2007). *Multiscale analysis of complex timeseries*. Hoboken, NJ: Wiley.
- Gilden, D. L. (2001). Cognitive emissions of 1/f noise. *Psychological Review*, 108, 33–56.
- Gilden, D. L. (2009). Global model analysis of cognitive variability. *Cognitive Science*, 33, 1–27.
- Holden, J. G. (2005). Gauging the fractal dimension of response times from cognitive tasks. In M. A. Riley, G. C. Van Orden (eds.), *Contemporary nonlinear methods for behavioral scientists: A webbook tutorial* (pp. 267–318). Retrieved August 23, 2009, from <http://www.nsf.gov/sbe/bcs/pac/nmbs/nmbs.jep>
- Holden, J. G., Choi, I., Amazeen, P. G., & Van Orden, G. (2011). Fractal 1/f dynamics suggest entanglement of measurement and human performance. *Journal of Experimental Psychology: Human Perception and Performance*, doi: 10.1037/a0020991
- Holden, J. G., Van Orden, G., & Turvey, M. T. (2009). Dispersion of response times reveals cognitive dynamics. *Psychological Review*, 116, 318–342.
- Ihlen, E. A., & Vereijken, B. (2010). Interaction-dominant dynamics in human cognition: beyond 1/f^α fluctuation. *Journal of Experimental Psychology: General*, 139, 436–463.
- Johnston, T., & DiTerlizzi, T. (2003). *Alien & Possum: Hanging Out*. New York: Simon & Schuster.
- Kantelhardt, J. W., Zschiegner, S. A., Koscielny-Bunde, E., Havlin, S., Bunde, A., & Stanley, H. E. (2002). Multifractal detrended fluctuation analysis of nonstationary timeseries. *Physica A*, 316, 87–114.
- Kiefer, A. W., Gresham, L. J., Wallot, S., Kloos, H., Riley, M. A., Shockley, K., & Van Orden, G. (2011). Development of coordination in time estimation. *Manuscript submitted for publication*.
- Kloos, H., & Van Orden, G. (2010). Voluntary behavior in cognitive and motor tasks. *Mind & Matter*, 8, 19–43.
- Marwan, N. (2011a). *Commandline recurrence plots*. Retrieved August 16, 2011, from <http://www.agnld.uni-potsdam.de/~marwan/6.download/rp.php>
- Marwan, N. (2011b). How to avoid potential pitfalls in recurrence plot based data analysis? *International Journal of Bifurcation and Chaos*, 21, 1003–1017.
- McAuley, J. D., Jones, M. R., Holub, S., Johnston, H. M., & Miller, N. S. (2006). The time of our lives: Lifespan development of timing and event tracking. *Journal of Experimental Psychology: General*, 135, 348–367.
- Michell, J. (1999). *Measurement in psychology: A critical history of a methodological concept*. Cambridge, UK: Cambridge University Press.
- Newell, A. (1973). You can't play 20 questions with nature and win: Projective comments on the papers of this symposium. In W. G. Chase (ed.), *Visual Information Processing* (pp. 283–308). New York: Academic Press.
- Peng, C.-K., Havlin, S., Stanley, H. E., & Goldberger, A. L. (1995). Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat timeseries. *Chaos*, 5, 82–87.
- Rayner, K., & Pollatsek, A. (1994). *The psychology of reading*. Englewood Cliffs, NJ: Prentice-Hall.
- Samuels, S. J. (1979). The method of repeated readings. *Reading Teacher*, 32, 403–408.
- Seuront, L. (2010). *Fractals and multifractals in ecology and aquatic science*. Boca Raton, FL: CRC Press.
- Shockley, K. (2005). Cross recurrence quantification of interpersonal postural activity. In M. A. Riley, G. C. Van Orden (eds.), *Tutorials in contemporary nonlinear methods for the behavioral sciences* (pp. 26–96). Retrieved August 23, 2009, from <http://www.nsf.gov/sbe/bcs/pac/nmbs/nmbs.jep>
- Shockley, K., Butwill, M., Zbilut, J., & Webber, C., (2002). Cross recurrence quantification of coupled oscillators. *Physics Letters A*, 305, 59–69.
- Riley, M. A., & Turvey, M. T. (2002). Variability and determinism in motor behavior. *Journal of Motor Behavior*, 34, 99–125.
- Takens, F. (1981). Detecting strange attractors in turbulence. In D. Rand, L.-S. Young (eds.), *Lecture Notes in Mathematics*, 898, 366–381.
- Van Orden, G. (2010). Voluntary performance. *Medicina*, 46, 681–594.
- Van Orden, G., Holden, J. G., & Turvey, M. T. (2003). Self-organization of cognitive performance. *Journal of Experimental Psychology: General*, 132, 331–350.
- Van Orden, G., Kello, C. T., & Holden, J. G. (2010). Situated behavior and the place of measurement in psychological theory. *Ecological Psychology*, 22, 24–43.
- Van Orden, G., Kloos, H., & Wallot, S. (2011). Living in the pink: Intentionality, wellbeing, and complexity. In C. A. Hooker (Ed.), *Philosophy of Complex Systems* (pp. 639–684). *Handbook of the Philosophy of Science* (Vol. 10). Amsterdam: Elsevier.
- Wallot, S., & Van Orden, G. (2011a). Grounding language performance in the anticipatory dynamics of the body. *Ecological Psychology*, 23, 1–28.

- Wallot, S., & Van Orden, G. (2011b). Toward a lifespan metric of reading fluency. *International Journal of Bifurcation and Chaos*, 21, 1173–1192.
- Wallot, S., & Van Orden, G. (2011c). Nonlinear analyses of self-paced reading. *The Mental Lexicon*, 6, 245–274.
- Webber, C. L. Jr., & Zbilut, J. P. (2005). Recurrence quantification analysis of nonlinear dynamical systems. In M. A. Riley, G. C. Van Orden (eds.), *Tutorials in contemporary nonlinear methods for the behavioral sciences* (pp. 26–96). Retrieved August 23, 2009, from <http://www.nsf.gov/sbe/bcs/pac/nmbs/nmbs.jep>.

Appendix: Software packages for recurrence and fractal analysis

We note a few online sources, where analysis tools for Fractal and recurrence analyses can be found because most of the analysis techniques presented in this chapter have not made it into the standard data analyses packages available to social scientists.

For monofractal analysis, Holden (2005) gives details instructions for how to construct a tool for standardized dispersion analysis, which is easily implemented in an Excel spread sheet (or any similar program). Also, the FracLab toolbox for Matlab has become popular and contains functions for mono and multifractal analysis. We found FracLab under <http://fraclab.saclay.inria.fr/>

A package for multifractal analysis, also for MatLab, is presently available on the website of the University of Trondheim:

<http://www.ntnu.edu/inm/geri/software>

In principle, monofractal structure can be assessed through multifractal analysis as well, since multifractal analysis is a generalization of monofractal analysis. The interested reader can find more about the topic in Gao, Cao, Tung and Hu (2007), Kantelhardt et al. (2002), or Ihlen and Vereijken (2010).

For recurrence analysis, quite a large selection of software packages can be found under:

<http://www.recurrence-plot.tk/programmes.php>

For Matlab users, the 'CRP toolbox' created by Norbert Marwan can be recommended. It possesses a good graphic interface, and provides the user with a visual image of the recurrence plot together with the statistics. Another reliable program is the program 'RQA software 13.1' of Charles Webber. The program is DOS based, however, and might require the installation of dos box or similar programs first.

Besides hosting recurrence analysis software, <http://www.recurrence-plot.tk/> boasts a nearly exhaustive bibliography of research that has employed recurrence quantification analysis, which can be an invaluable resource.

Brain imaging and conceptions of the lexicon

Jed A. Meltzer

Rotman Research Institute — Baycrest Centre

Neuroimaging plays an increasingly important role in the investigation of all aspects of human cognition, including language. Historically, experimental psychology and neuroimaging relied on very different techniques, as neuroimaging studies required comparisons between different tasks rather than manipulation of conditions within a single task, as is standard in behavioural experiments. However, methodology has advanced in the past decade such that many classic behavioural paradigms can now be employed in studies that measure brain activity. We review the technical foundations of conducting studies on single-trial brain responses, using event-related fMRI and electrophysiological recordings. We focus in particular on studies of picture naming, illustrating how the same techniques that were originally used to define temporal processing stages in reaction time studies can now be applied to brain imaging studies to reveal the neural localization of those stages.

Introduction

The explosive growth of neuroimaging as a scientific paradigm has attracted intense enthusiasm along with criticism. While philosophical objections to the “phrenological” nature of modern cognitive neuroscience are nothing new, even those researchers who accept the principal of localization of function can find plenty of reasons to be skeptical about neuroimaging findings. For most of its history, the entire enterprise of functional neuroimaging has been critically dependent on the so-called “logic of cognitive subtraction.” In its most basic form, this paradigm involves testing for the neural localization of a putative cognitive function by measuring brain activity during two contrasting tasks: an “active task” that includes the function of interest, and a “baseline task” that does not, but is well-matched on other factors.

While the subtractive paradigm has its advantages and disadvantages, one of the major obstacles to the acceptance of functional imaging in mainstream experimental psychology may be the fundamental differences in experimental design in