The Fractal Fabric of Speech

Gregory G. Anderson (ganders5@gmu.edu)
George Mason University, 4400 University Drive
Fairfax, VA, 22030-4444, USA

Christopher T. Kello (ckello@gmu.edu)
George Mason University, 4400 University Drive
Fairfax, VA, 22030-4444, USA

Abstract

An experiment was conducted in order to characterize the intrinsic fluctuations of human behavior as they are reflected in multiple repetitions of a single spoken word. Ten participants repeated the word "bucket" 1100 times, and fluctuations across repetitions in the acoustic measures of syllable duration, peak pitch, peak intensity and spectral intensity were analyzed for power law scaling relations. All measures for all subjects showed fluctuations resembling the scaling relation known as *1/f noise*, with many distinct streams of 1/f noise running in parallel. These results provide evidence for the emergent basis of human behavior.

Introduction

At some level, everyone would agree that an individual's behavior is the product of many neural and bodily systems working together, influenced broadly by the individual's historic and behavioral contexts. Our question is "how do these systems coordinate to produce coherent behavior?" Coherence is ubiquitous to human behavior but can be seen clearly in transparent examples like swimming or drumming, where the limbs exhibit a coherent orchestration of movement, presumably with a corresponding orchestration of neural activity.

The implications of this question about the fundamental basis of coordination can be illustrated by considering the interpretation of neuroimaging results. Many studies report that multiple brain regions are engaged in performing cognitive tasks (Cabeza & Nyberg, 2000), and it is a general rule that activation becomes more widespread as task difficulty increases. It must also be recognized that the amount of activation observed in neuroimages is just as much a function of analysis parameters as it is of actual neural activity. Therefore it appears that activation of multiple brain regions can be observed for any given behavioral performance.

Evidence for widespread brain activity raises the question of how brain regions interact to produce behavior. If their interactions are linear or weakly non-linear, then one could plausibly use neuroimaging to identify the contributions that each region makes to behavior, abstracted away from the particulars of behavioral contexts. In other words, one could plausibly draw causal lines from brain regions to behavioral categories. As just one example from language research, one could use subtractive or correlational methods (which are linear) to test whether one set of regions is used in reading words, and a different set of regions is used in reading nonwords.

But if interactions are sufficiently non-linear, then the roles of brain regions in behaviors will be strongly conditioned by context (e.g., see Elman et al., 1996). To provide an illustrative analogy using the human body, consider how the roles of the hands are conditioned by an individual's history and behavioral context. They can be used for grasping, chopping, gesturing, sign language, playing music, and even walking, under the right conditions. While one might try to fix the thread that runs through these functions, it would be very difficult to capture the seemingly boundless number of *potential* functions of the hands. This space of potential functions is created by context, which may include the task demands as well as many aspects of an individual's body and experience.

The analogy is that, if interactions are sufficiently non-linear, then the functions of brain regions are contextual in the same way that the functions of the hands are contextual. This does not mean that each brain region is equally likely to support any given function. To the contrary, brain regions undoubtedly have functional distinctions, just as the hands are functionally distinct from other anatomical components. But these distinctions would derive from differences in the *potentials and constraints* that are provided by each component to shape the contextual *emergence* of cognitive and behavioral function.

We use the term emergence to mean that function is instantiated by coordinations of activity whose cause cannot be isolated in one or more system components. Instead, the cause of coordination lives exclusively "in between" the components, i.e., in their two-way interactions. Such emergent coordination is defining of *complex systems*, and relatively simple models of emergent coordination have been studied in complex systems throughout the physical, biological, and social science (see Bak, 1996; Holland, 1998; Kelso, 1997).

Our working hypothesis is that human systems are complex systems, in the sense that their components are governed by non-linear interactions from which the coordination of behavior emerges. This is a very general hypothesis that one might not at first expect to be empirically testable, but it turns out that the capacity for emergent coordination in complex systems has a universal signature. From black holes to quasars, rivers to fault lines, financial networks to computer networks and brains to hearts, these and many other kinds of complex systems have all exhibited *power law scaling relations* in their behaviors. Scaling relations in nature are generally accepted as emergent patterns of coordination (West & Brown, 2005).

Scaling relations occur most generally when one system variable is related to another system variable raised to some power. Scaling relations are *fractal*, in that the variables bear a self-similar relation to each other across scales of measurement (Bak, 1996). To illustrate, let us introduce the scaling relation known as *long-range correlation* (Wagenmakers, Farrell, & Ratcliff, 2005). The experiment reported herein investigates long-range *temporal* correlations in human behavior, but it is instructive to first go through an example of long-range *spatial* correlations.

Consider a sheet of cortex as a system of neurons. Imagine flattening the sheet and defining the distance between two neurons as their Euclidean distance on the sheet. Further imagine that each neuron has a time-varying level of activity, and that correlations in neuronal activity are measured as a function of distance apart. This unrealistic but illustrative model of cortex would exhibit long-range spatial correlations if the activities of nearby neurons were positively correlated, and correlations decayed towards zero slowly as distance increased. Slow decay means specifically that correlations diminish as an *inverse power* of distance, rather than the exponential decay that is more commonly discussed (exponential decay corresponds to "short-range" correlation). In particular, $C(d) \approx 1/d^{\gamma}$, where $C(d)$ is correlation as a function of distance d, and $0 < \gamma < 1$.

Power law decay of correlations means that all neuronal activities would tend to be correlated with each other to some degree, no matter how far apart (hence the term "long-range"). Long-range spatial correlations indicate a coherence of activity across the entire model sheet of cortex, and hence the potential for emergent coordination. This potential has been demonstrated by simple models in which long-range spatial correlations lead to spontaneous, global patterns of neural activity (Stauffer & Aharony, 1992).

Long-range temporal correlations are defined by the same power law decay, but for measurements across different points in time rather than different points in space. In our model sheet of cortex, for instance, one can imagine measuring the activity of one neuron over two different time periods, and testing whether the two time series of measurements are correlated. The neuron would exhibit long-range temporal correlations if $C(k) \approx 1/k^{\gamma}$, where k is defined in terms of separation in time instead of space. Analogous to our spatial illustration, long-range temporal correlations may indicate the potential for coordinated activity to emerge and cohere across time.

If one accepts long-range temporal correlations as evidence for the emergent basis of human behavioral coordination, then the evidence has been mounting for some time (Gilden, 2001; Van Orden, Holden, & Turvey, 2003). Long-range temporal correlations have been found in many kinds of human performances, including walking, gazing, finger tapping, and ratings of self-esteem (see Van Orden et al., 2003). Moreover, these long-range correlations have been found to resemble the specific power law known as *1/f noise*.

To illustrate the properties of 1/f noise, a time series of 1024 simple reaction times is shown in the left panel of Figure 1, taken from one participant in an experiment reported by Beltz and Kello (2004). Distinct undulations can be seen in the time series, undulations that extend across dozens and even hundreds of reaction times. These undulations have a fractal or self-similar quality, in that their statistics are the same across timescales. In other words, one could "zoom in" or "zoom out" on the time series to find undulations nested within undulations, such that one would not be able to determine the scale of measurement based on the statistics of the visible undulations. 1/f noise is thus said to be *scale-free*.

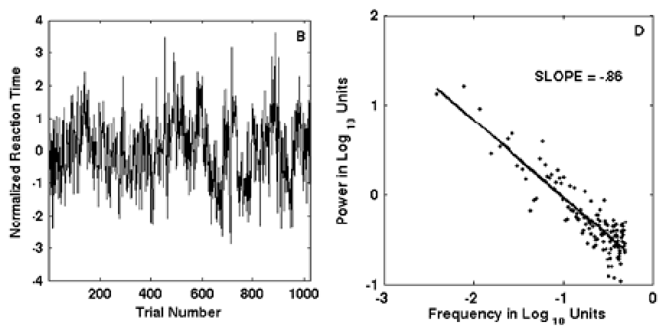

Figure 1: Left graph shows reaction times plotted as a trial series. Right graph shows a spectral analysis of the time series, plotted in log-log coordinates.

The 1/f scaling relation can been seen more clearly in a spectral analysis of the time series, shown in the right panel of Figure 1. Spectral analysis essentially decomposes the time series into a set of sine waves of varying amplitudes and frequencies. Each point on the spectral plot represents one sine wave, with its frequency on the x-axis and power (squared amplitude) on the y-axis. Scaling relations express themselves as linear trends in log-log coordinates, and the regression line in Figure 1 indicates a clear scaling relation between power and frequency in this participant's reaction time fluctuations. In particular, $P \approx 1/f^{\alpha}$, where alpha is estimated in the range $0 < \alpha < 1$.

This range is noteworthy because 1/f noise with α near one has been found throughout complexity phenomena in nature (for hundreds of examples, see http://www.nslij-genetics.org/wli/1fnoise). Moreover 1/f noise strikes a balance between the randomness of white noise (where α is near zero) and the regularity of brown noise (i.e., a random walk were α is near two). Despite decades of research on 1/f noise, this scaling relation has proven difficult to interpret because many models are known to generate or mimic the basic finding of 1/f noise, and not all of them are models of emergence as we have defined it. Thus a skeptic may say that 1/f noise is ubiquitous simply because there are so many unrelated ways for it to occur. The same could be said for nearly all of the reports of 1/f noise in human performance to date.

Current Experiment

We designed an experiment to provide more discriminating evidence on whether 1/f noise is a sign of the emergent basis of coordination in human behavior. The basic logic is that, if 1/f noise is a generic property of system interactions, then it should be found wherever the collective effects of system interactions are measured *without being obscured by task-specific effects*.

This qualification turns out to be a rather stringent one. The issue is that, for any given series of measurement trials, idiosyncratic effects will be introduced nearly anytime that behavior is purposely varied in some way from one trial to the next. Such variations are unwanted because they will affect behavioral measurements according to the order in which they occur, and these order effects will be reflected in the spectral portraits of behavioral fluctuations. Therefore, in order to observe 1/f noise most clearly, one should take measurements of the same behavior enacted repeatedly, with minimal perturbations and contingencies from one enactment to the next. The only distinction between measurements should be that they occurred at different points in time. We shall refer to such measurements as the *intrinsic fluctuations* of behavior.

1/f noise in human behavior has thus far been consistently reported in measurement conditions that approximate to some degree the pure definition of intrinsic fluctuation. For instance, one of the clearest and earliest reports of 1/f noise in human behavior came from the task of repeatedly estimating the same distance or amount of time, over and over again with no external cue for more than 1000 times (Gilden, Thornton, & Mallon, 1995) (It is general practice that evidence for a scaling relation needs to span at least three decades of scale, which requires over 1000 data points). More explicitly, Beltz and Kello (2004) found that behavioral fluctuations were de-correlated by perturbations to measurement in the form of unpredictable variations in a cue to respond.

Results to date are consistent with the concept of intrinsic fluctuation and the general hypothesis of emergent coordination, but they may still be explained by non-emergent hypotheses. For instance, under certain parameterizations, 1/f noise may result from the summation of processes that fluctuate on a wide range of timescales (Beran, 1994). Perhaps variations in the timescales of bodily and neural processes happen to align to produce 1/f noise (Bills, 1935), or variations in unconscious, subconscious, conscious processing may similarly align (Ward, 2002). Another possibility is that attentional or strategic drifts might follow a pattern of 1/f noise for some reason (Pressing & Jolley-Rogers, 1997; Wagenmakers, Farrell, & Ratcliff, 2005). These non-emergent mechanisms may also explain the association of 1/f noise with intrinsic fluctuation, but without a commitment to emergence.

The non-emergent basis of these alternate mechanisms leads to a testable prediction. The prediction is based on the fact that the alternate mechanisms are all singular, isolated sources of 1/f noise. At any given point in time, they predict only one "signal" of 1/f noise to be emitted from a person. There can be only one overall "system flux", for instance, and multiple threads of attention or strategy are not typically hypothesized. By contrast, according to emergent coordination, 1/f noise is a generic property of system interactions that give rise to all behaviors. Any and all behavioral signals should yield 1/f noise under conditions of intrinsic fluctuation, even if there are multiple distinct signals. Thus one should be able to find multiple, parallel streams of 1/f noise under conditions of intrinsic fluctuation.

Beltz and Kello (2004) tested these competing predictions by creating conditions for measuring two parallel but unrelated streams of intrinsic fluctuation. As already mentioned, they measured fluctuations in reaction times to simple response cues, but they also measured fluctuations in the corresponding *key-contact durations*, that is, the time from key press to key release. Reaction times and key-contact durations were indeed uncorrelated as one might expect, yet they both fluctuated as 1/f noise.

While the appearance of these two parallel yet distinct streams of 1/f noise was predicted by emergent coordination, the result can be accommodated post-hoc by non-emergent hypotheses. For instance, one might claim that conscious/controlled/cognitive processes emit their own stream of 1/f noise in reaction times, and unconscious/automatic/motor processes emit a second stream of 1/f noise in key-contact durations. Kello and his colleagues (2006) argued against this kind of post-hoc account on logical grounds, but such accounts remain as logical possibilities.

The experiment herein was designed to push this issue to its logical extreme. If 1/f noise reflects the emergent basis of behavioral coordination, then the possible number of parallel and distinct streams of 1/f noise should be *unlimited* in principle. Key presses afford no more than a small handful of behavioral measures, and so are ill-suited to testing for many parallel and distinct streams of 1/f noise. Therefore, we elicited intrinsic fluctuations in a repeated speech token because spoken utterances afford a plethora of dependent measures to examine for 1/f noise. If 1/f noise reflects the emergent basis of behavioral coordination, then all acoustic measures of intrinsic fluctuation, no matter how distinct from each other, should all appear as 1/f noise.

Methods

Participants. Five male and five female undergraduate students participated in the experiment for course credit**.**

Procedure. Each participant said the word "bucket" 1100 times in a row. Utterances were paced by audiovisual cue that was presented once every 1.2 seconds. The word "bucket" was chosen because it is easy to produce and it affords the automation of many acoustic measures. Participants were fitted with headworn cardioid microphones to reduce perturbations of the recorded signal from environmental noises and changes in microphone proximity. They were instructed to speak in a natural manner and to produce one "bucket" after each cue. A practice block of ten utterances preceded the experimental block.

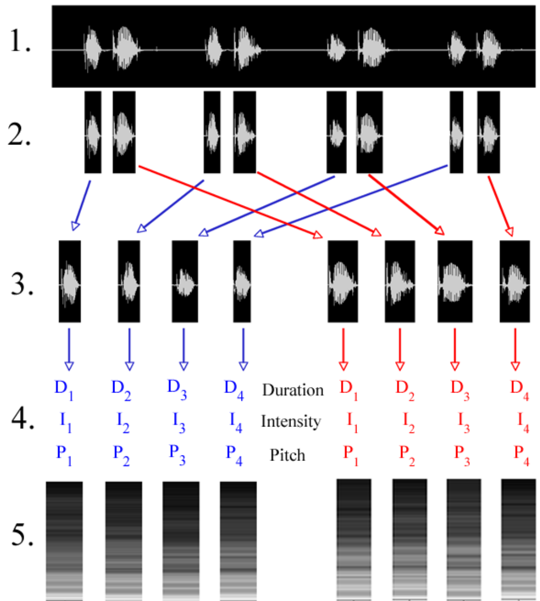

Figure 2: Steps in data collection process: 1) Audio file of block of utterances; 2) Segmented audio files by syllable; 3) Serial order by syllable; 4) Duration, intensity and pitch values determined for each syllable; 5) Spectrograph-like plots of spectral power estimates, coded black (low power) to white (high power). Ordered from low frequency (bottom) to high frequency (top).

Data Collection and Analysis. Each utterance was segmented into separate syllables using standard automated tools and parameters that are part of the Logic Pro digital audio software. Two non-spectral acoustic measurements were taken from each syllable using the Praat speech analysis software (Boersma & Weenink, 2005): peak intensity and acoustic duration. Peak pitch was also measured from the "buck" syllables but not the "ket" syllables because the latter generally did not contain enough periodic energy. In addition to these five non-spectral measures per utterance, the long-term average spectrum (LTAS) was computed for each syllable. The LTAS was analyzed using frequency bands that were 160 Hz wide, with center frequencies from 80Hz to 13,440Hz, yielding a total of 84 spectral power estimates per syllable. In all, 173 acoustic measurements were taken per utterance (see Figure 2). Utterances with artifacts or anomalous measurements were removed from the analyses, and then the beginnings of each data series were truncated to yield 1024 utterances per participant.

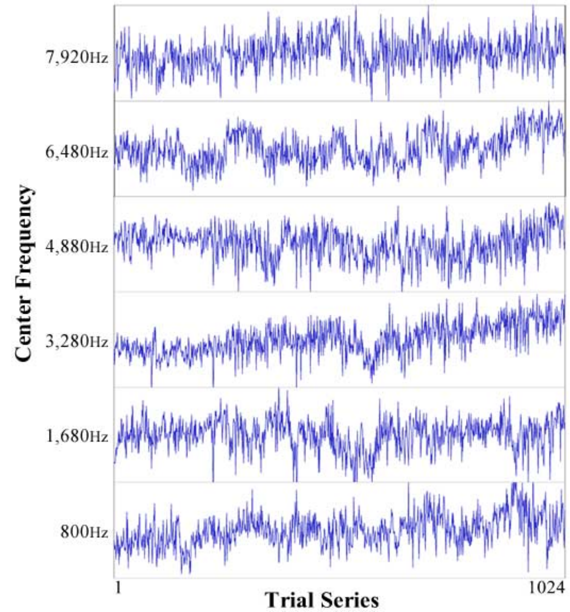

Figure 3: Examples of spectral power fluctuations for one participant over the course of the experiment.

Results

Figure 3 shows six example time series of spectral power estimates for one participant's series of "buck" syllables. Two informal observations can be made: each series exhibits the self-similar nested undulations that are characteristic of 1/f noise, and the series do not appear to be positively correlated with each other.

To more formally examine the data for 1/f noise, a spectral analysis (not to be confused with the LTAS measures) was conducted on the fluctuations across utterances for each of the 173 acoustic measures taken from each participant. The resulting spectra for each acoustic measure were then averaged across participants, and the averaged spectra for the non-spectral measures are shown in Figure 4 in log-log coordinates. Each individual spectrum resembled the average, and the averages show clear scaling relations, particularly in the lower frequencies. The higher frequencies are less reliable because they are more greatly influenced by measurement error, which "whitens" the spectral portrait. Whitening is seen as a flattening of the 1/f scaling relation.

Therefore, to estimate the exponent of the scaling relation seen in the averaged data, regression lines were fit to the lowest 32 frequency bins of each spectrum. The slopes of these lines were used as negative estimates of the exponent of the $1/f^{(}$ scaling relation. While one may debate the strengths and weaknesses of this estimation method compared with others (Thornton & Gilden, 2005), all methods are likely converge to similar estimates in this case, given the clarity of the spectral portraits for these data. Furthermore, we are only trying to estimate whether the exponent falls somewhere within the range of 1/f noise, rather than pinpointing its specific value.

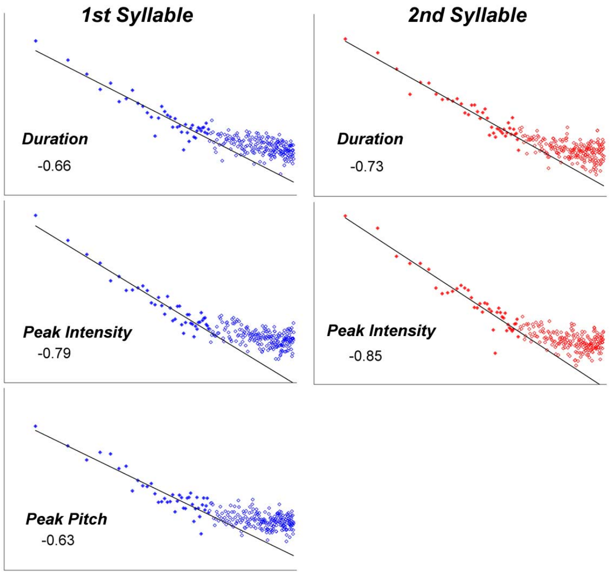

Figure 4: Power spectra averaged across participants for acoustic duration, peak intensity and peak pitch. Log frequency is on the x-axis and log power is on the y-axis. Negative estimates of the 1/f$^{\langle}$ exponents are shown.

The next question to ask of these data is whether they reflect multiple, distinct streams of 1/f noise. To test this, we computed correlation coefficients for all pairs of measures for each participant, and the average coefficients ranged from 0.05 to 0.22. The lack of correlation among these acoustic measures is clear evidence for as many as five distinct streams of 1/f noise.

The same analyses were also conducted on the 168 LTAS measures of intrinsic fluctuation, 84 per syllable. The results of these analyses are presented in Figures 5 and 6. In the top half of Figure 5, the negative estimates of the exponents (i.e., regression line slopes) for averaged spectra are plotted as a function of frequency for each syllable. In the bottom half of Figure 5, the spectra are averaged across frequency for each syllable. These graphs show that the 1/f scaling relation was ubiquitous to all the spectral measures of intrinsic fluctuation in the repetitions of "bucket". Figure 6 shows the spectral analyses for each participant separately, averaged across the frequency bins. These graphs show that all ten participants generated the same 1/f scaling relation, with the exponent ranging from 0.50 to 0.87 across participants.

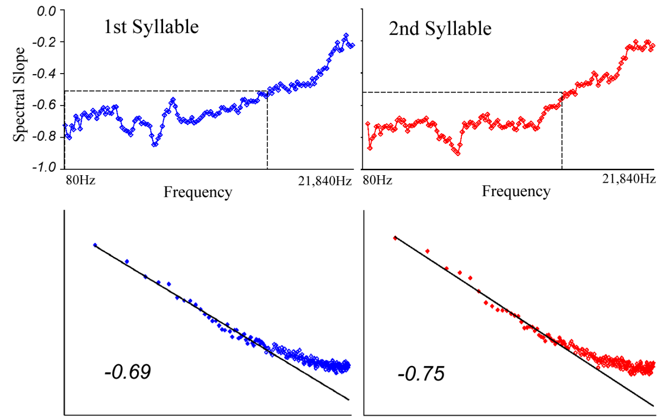

Figure 5: Top graphs plot the spectral slope estimates as a function of frequency band for data averaged across participants. The frequency range for speech is shown by the dotted lines. Bottom graphs plot the log-log spectra averaged across participants and frequency bins.

Finally, to test whether the LTAS intrinsic fluctuations contained distinct streams of 1/f noise, a 168 by 168 correlation matrix was computed for each participant's data, and the resulting correlation matrices were averaged together. Inspection of the averaged matrix showed hundreds of pairs of uncorrelated LTAS components (i.e., average coefficients of ±0.1). Principal components analysis (PCA) was then used to derive a more specific estimate of the number of purely uncorrelated (orthogonal) streams of 1/f noise. The 84 LTAS series for each syllable were submitted to PCA analysis, and the resulting orthogonal fluctuations (i.e., the original data projected onto the principal components) were submitted to spectral analysis to retest for the 1/f scaling relation.

The results showed that the strongest components of the data also had exponent estimates closest to the boundary condition of one for 1/f noise. Moreover, dozens of components for each syllable, all uncorrelated by definition, were estimated to be well within the range of 1/f noise. For instance, 11 components in the first syllable and 14 components in the second syllable had exponents estimates

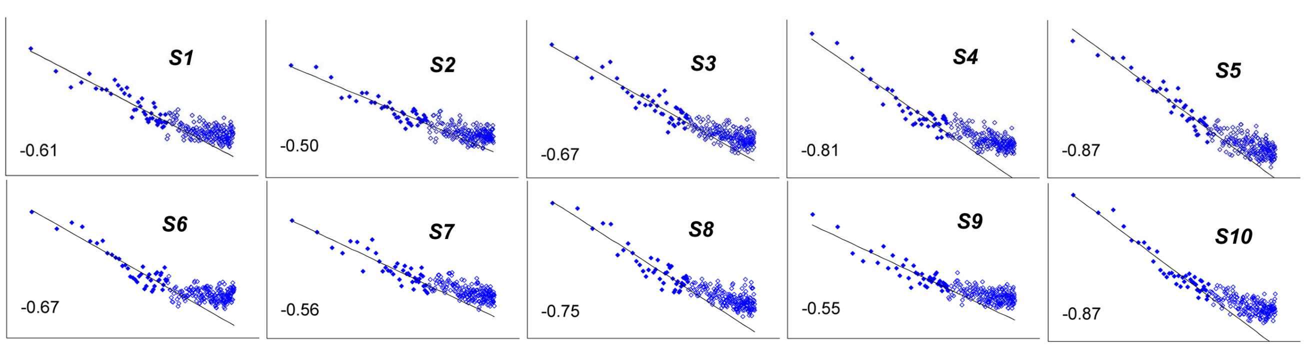

Figure 6: Spectral analyses of the LTAS fluctuations, averaged across frequency bins and separated by participant.

greater than 0.5. These components accounted for well over 90% of the data. The remaining components fluctuated as white noise.

Conclusion

The specific aim of our experiment was to test whether multiple, parallel streams of 1/f noise can be found in a rich and intricate human behavior like speech. Over 100 acoustic measures of the word "bucket" were observed under conditions of intrinsic fluctuation, and all measures for all ten participants closely resembled the power law scaling relation known as 1/f noise. Moreover, correlations and PCA analyses indicated that dozens of distinct streams of 1/f noise ran in parallel through the intrinsic fluctuations of speech.

These findings were predicted by the hypothesis that 1/f noise originates from interactions that are generic to human systems, and that provide for the emergent basis of human behavior. By contrast, the findings are difficult to reconcile with the hypotheses that 1/f noise originates from system flux, or fluctuations in attention, strategies, and the like. The problem with these non-emergent hypotheses is that they must proliferate post-hoc mechanisms for every distinct stream of 1/f noise that is observed.

If one accepts these results along with many other studies as evidence for the emergent basis of human behavior, then the next step is to more specifically characterize the system interactions that give rise to 1/f noise under conditions of intrinsic fluctuation, and more generally give rise to the coordination and coherence of human behavior. Similar issues of emergent coordination have been studied rigorously in the physical sciences for decades. The finding of power law scaling relations in the cognitive sciences opens the way for physical and biological models to be used more extensively as sources of inspiration for theories of human cognition and behavior.

Acknowledgements

This work was funded in part by NSF Grant 0239595.

References

Bak, P. (1996). How nature works : the science of self-organized criticality. New York, NY: Copernicus.

Beltz, B. C., & Kello, C. T. (2004). The effect of cue predictability on long-range dependence in response times versus response durations. Paper presented at the *Proceedings of the 26th Annual Meeting of the Cognitive Science Society*.

Beran, J. (1994). *Statistics for long-memory processes*. New York: Chapman & Hall.

Bills, A. G. (1935). Fatigue, oscillation and blocks. *Journal of Experimental Psychology, 18*, 562-573.

Boersma, P., & Weenink, D. (2005). *Praat: doing phonetics by computer* (Version 4.3.30).

Cabeza, R., & Nyberg, L. (2000). Imaging cognition II: An empirical review of 275 PET and fMRI studies. *Journal of Cognitive Neuroscience, 12*, 1-47.

Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness : a connectionist perspective on development*. Cambridge, Mass.: MIT Press.

Gilden, D. L. (2001). Cognitive emissions of 1/f noise. *Psychological Review, 108*, 33-56.

Gilden, D. L., Thornton, T., & Mallon, M. W. (1995). 1/f noise in human cognition. *Science, 267*(5205), 1837-1839.

Holland, J. H. (1998). *Emergence : from chaos to order*. Reading, Mass.: Addison-Wesley.

Kello, C. T., Beltz, B. C., Holden, J. G., & Van Orden, G. C. (2006). Fractal Streams in Human Performance. *Manuscript under revision*.

Kelso, J. A. S. (1997). *Dynamic patterns : the self-organization of brain and behavior*. MIT Press.

Pressing, J., & Jolley-Rogers, G. (1997). Spectral properties of human cognition and skill. *Biological Cybernetics, 76*, 339-347.

Stauffer, D., & Aharony, A. (1992). *Introduction to percolation theory* (2nd ed.). Washington, DC: Taylor & Francis.

Thornton, T. L., & Gilden, D. L. (2005). Provenance of correlations in psychological data. *Psychonomic Bulletin & Review, 12*, 409-441.

Van Orden, G. C., Holden, J. G., & Turvey, M. T. (2003). Self-organization of cognitive performance. *Journal of Experimental Psychology. General, 132*(3), 331-350.

Wagenmakers, E. J., Farrell, S., & Ratcliff, R. (2005). Human cognition and a pile of sand: a discussion on serial correlations and self-organized criticality. *Journal of Experimental Psychology: General, 134*(1), 108-116.

Ward, L. M. (2002). *Dynamical cognitive science*. Cambridge, Mass.: MIT Press.

West, G. B., & Brown, J. H. (2005). The origin of allometric scaling laws in biology from genomes to ecosystems: towards a quantitative unifying theory of biological structure and organization. *Journal of Experimental Biology, 208*, 1575-1592.