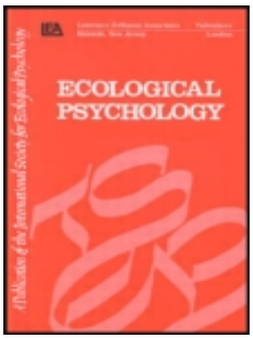

Ecological Psychology

ISSN: 1040-7413 (Print) 1532-6969 (Online) Journal homepage: www.tandfonline.com/journals/heco20

Production and Convergence of Multiscale Clustering in Speech

Drew H. Abney, Christopher T. Kello & Anne S. Warlaumont

To cite this article: Drew H. Abney, Christopher T. Kello & Anne S. Warlaumont (2015) Production and Convergence of Multiscale Clustering in Speech, Ecological Psychology, 27:3, 222-235, DOI: 10.1080/10407413.2015.1068653

To link to this article: https://doi.org/10.1080/10407413.2015.1068653

Ecological Psychology, 27:222–235, 2015

ISSN: 1040-7413 print/1532-6969 online
DOI: 10.1080/10407413.2015.1068653

Routledge
Taylor & Francis Group

Production and Convergence of Multiscale Clustering in Speech

Drew H. Abney, Christopher T. Kello, and Anne S. Warlaumont
Cognitive and Information Sciences
University of California, Merced

Language entails the scaling of variability across levels of measurement—small linguistic variations occur at the millisecond level, larger variations occur at the next level, and even larger variations occur over longer timescales. For acoustic onsets in speech signals, small temporal variations occur at the phonetic level, larger variations occur at the phrasal level, and even larger variations occur at the conversational level. Scaling across levels of measurement can be quantified in terms of power law distributions. In this article we review recent investigations into power law clustering of acoustic speech onsets. Studies demonstrate that the multiscale clustering in onsets reflects communicative aspects of speech in adult conversations as well as infant vocalizations. We also review evidence that multiscale clustering in the vocalizations of individuals converges during vocal interactions. We relate multiscale convergence to the notion of complexity matching, that is, the hypothesis that maximal information transfer occurs when the power laws of 2 interacting complex systems are matched. We conclude by discussing potential extensions of this work including estimating the multifractal structure of speech and testing the maximal information transfer prediction of complexity matching.

Language displays hierarchically nested structures: phonemes are nested in syllables, syllables in words, words in phrases, phrases in sentences, and sentences in discourse. One consequence of this hierarchy is that the variability

Correspondence should be addressed to Drew H. Abney, Cognitive and Information Sciences, University of California, Merced, School of Social Sciences, Humanities and Arts, 5200 North Lake Road, Merced, CA 95343. E-mail: drewabney@gmail.com

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/heco.

within the system scales across levels of measurement (Bak, 1996; Bassingthwaighte, Liebovitch, & West, 2013; Kello, Beltz, Holden, & Van Orden, 2007; Kelso, 1997; Van Orden, Holden, & Turvey, 2003). Consider the variability in timing of acoustic onsets during speech production: small variations occur in small clusters of onsets over tens of milliseconds, larger variations in larger clusters spanning hundreds of milliseconds, and even larger variations occur over minutes and longer periods of time. Variability of measured behavior that scales across levels of measurement is indicative of a type of nonlinear relation, a power law, and such power laws emerge for systems exhibiting hierarchically nested structures like language (Mandelbrot, 1983).

In this article, we discuss the multiscale patterns of vocalization variability that humans produce when using language in conversational speech and also in the vocalizations produced by infants and caregivers. In doing so, we emphasize the idea that hierarchical structure in language can be expressed as the temporal clustering in speech across multiple temporal levels of analysis, where temporal clustering is measured in terms of the timing of acoustic onsets. The sections that follow emphasize the degree to which clustering in acoustic onsets grows with timescale, a term we call *multiscale clustering*, in human vocalizations. We discuss how multiscale clustering might emerge and advance through development and how multiscale clustering of vocalizations converges during vocal interactions, as theorized and measured by complexity matching.

MULTISCALE CLUSTERING IN SPEECH ACOUSTICS

Because language has temporally nested organization, such as phonetic boundaries within phrasal boundaries within conversational turns (Pickering & Garrod, 2004), there is good reason to also expect multiscale clustering of acoustic onsets in conversational speech. For example, Figure 1 illustrates how amplitude-based acoustic onsets from one interlocutor's speech (black vertical lines above each of the three waveforms) can cluster across multiple timescales during conversation. Notice that there are clusters of acoustic onsets at all three scales, that is, sets of onsets that occur in close temporal proximity relative to a given timescale. These sets may consist of only a handful of onsets at small timescales or dozens or hundreds of onsets at longer timescales. For instance, the top waveform shows a 10-min span of speech. A cluster may consist of hundreds of onsets at this timescale reflecting a speaking turn. The middle waveform shows a 2-min span of speech. A cluster may consist of dozens of onsets at this timescale reflecting an utterance. Finally, the bottom waveform shows a 20-s span of speech. A cluster may consist of 10 or 20 onsets at this timescale reflecting a sentence or a phrase. At even smaller timescales (not pictured) we could see clusters of onsets reflecting temporal patterning of syllables.

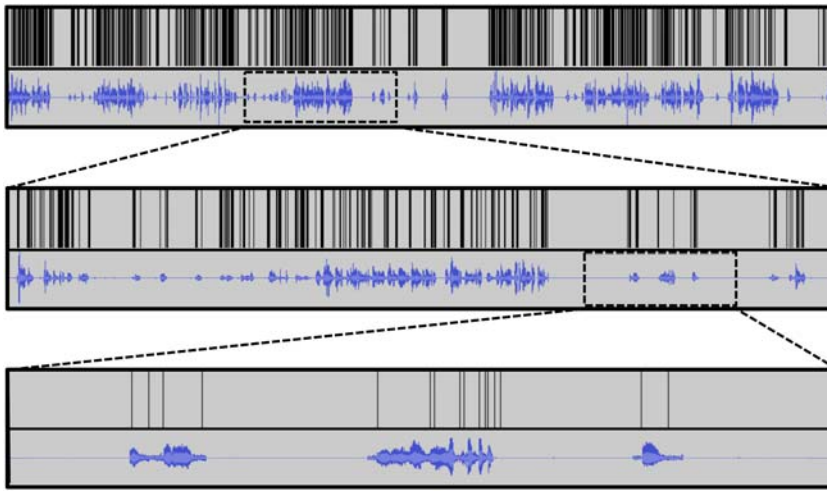

FIGURE 1 An example conversational speech signal shown at three different temporal scales, 10 min (top), 2 min (middle), and 20 s (bottom). Figure adapted from Abney, Paxton, Dale, and Kello (2014), *Journal of Experimental Psychology: General*.

We can measure if, and to what degree, temporal clustering of onsets exists at multiple timescales by estimating whether clustering is related to timescale by a power law function (see Clauset, Shalizi, & Newman, 2009). A power law function expresses one variable as a nonlinear function of another variable raised to a power, for example, $A(T) \sim T^{\alpha}$. In this case, in correspondence with the three timescales depicted in Figure 1, T is a particular timescale (e.g., 20 s, 2 m, 10 m) and $A(T)$ is the coefficient of variability of acoustic events at that particular timescale. The scaling exponent, α, can be determined by plotting the timescales on a logged x-axis and coefficient of variability on a logged y-axis and then estimating the slope from a regression line. If the function observed is linear, the slope corresponds to α. It should be noted that most power law analyses systematically increase the window size by successive exponents with a constant base, for example, 2^1, 2^2, 2^3, 2^4, and so on. Different types of power law functions vary along many dimensions such as the unit of analysis (e.g., event-based or interval-based) and the type of scale (e.g., temporal scale or a spatial scale). These dimensions can constrain the properties of a phenomenon researchers might target. A suitable analysis is required to measure the temporal clustering in speech across multiple temporal scales. The Allan Factor (AF) analysis (Allan, 1966) has been recently applied as a tool for measuring if, and to what degree, a speaker's time series of acoustic events exhibited multiscale clustering.

We now provide an informal description of the AF analysis. For the interested reader, we have included a formal description of the AF analysis in the Appendix. The AF analysis takes a binary spike train and estimates event-based variation over a range of temporal windows (i.e., timescales). For our purposes, the binary spike train consists of 0s and 1s, where 1s are when acoustic onsets occurred in the event series. The AF estimates are then logarithmically plotted on the y-axis and the corresponding temporal window is logarithmically plotted on the x-axis. Multiscale clustering of acoustic onsets is exhibited if an estimated slope of a regression line, α, is greater than $\alpha \sim 0$. The null hypothesis for the AF analysis is that the slope is $\alpha \sim 0$, which would indicate that the clustering of acoustic onsets does not scale across timescales. In other words, an α that is greater than ~ 0 is indicative of an event series with variability that scales over time, whereas an $\alpha \sim 0$ is indicative of an event series with variability that does not scale over time. More formally, the sequence of acoustic onsets is a fractal point process if $0 < \alpha < 3$ and a homogeneous Poisson process if $\alpha \sim 0$ (Lowen & Teich, 1996). AF is the event-based analog (Viswanathan, Peng, Stanley, & Goldberger, 1997) to spectral analysis and detrended fluctuation analysis (DFA; Peng, Havlin, Stanley, & Goldberger, 1995). Spectral and DFA analyses are the standard methods for measuring long-range correlations in cognitive and behavioral processes.

Recent work in our lab found evidence of multiscale clustering in conversational speech signals (Abney, Paxton, Dale, & Kello, 2014) from reanalyzing data from Paxton and Dale (2013). In Paxton and Dale, two adult participants were instructed to have one 10-min conversation about their favorite movies, music, books (affiliative conversation), and one 10-min conversation about a controversial topic on which they had opposing opinions (e.g., opinions about the death penalty; argumentative conversation). During both types of conversation, the participants produced patterns of acoustic energy that clustered across temporal scales (see Figure 1). We used AF analysis to measure the variation of event clustering of acoustic onsets, AF(T), across multiple timescales, T. The smallest timescale was $T = 160$ ms, and the largest timescale was $T = 41$ s. We found that the variation of temporal clustering of acoustic onsets scaled across timescales (160 ms to 41 s) that approximate levels of linguistic representation such as phonetic, lexical, semantic, and discourse levels (Pickering & Garrod, 2004). Notably, we also found that multiscale clustering of vocalizations differed as a function of the type of conversation two people were having. The scaling exponent, α, was larger for argumentative ($\alpha = .63$) than for affiliative ($\alpha = .53$) conversations, indicating more multiscale clustering, in the speech data, for the argumentative conversations. These results suggest that the onset and clustering of low-level acoustic energy are constrained by, and sensitive to, linguistic context. These results also suggest that in general the onset and clustering of conversational speech is non-Poisson and fractal in nature.

Investigating the multiscale dynamics of speech by estimating the multiscale clustering of acoustic onsets is a relatively new research direction. However, previous research with similar intuitions about the multiscale dynamics of speech and language focused on the coupling of multiple frequencies of speech production as oscillatory dynamical systems. Cummins and Port (1998) used a phrase repetition task and observed phase coupling of oscillators at the levels of phrase and foot. Tilsen (2009) observed rhythmic-gestural covariability across gestural units and prosodic units like phrase, foot, and syllable. These studies focused on specific rhythmic units, perceptual centers or p-centers, which are beatlike events estimated at the halfway point of the sonority rise toward a nuclear vowel. From p-centers, rhythmic oscillatory models can be constructed.

To date, the unit of measurement for temporal clustering has been the onset of an acoustic event defined either by an amplitude threshold, pitch, or a combination of these properties. We have seen similar scaling patterns across acoustic events derived from all of these properties. The fact that we observed similar scaling patterns in our AF analyses for various types of acoustic events indicates that the observed multiscale clustering may not be too sensitive to how vocalization onset is defined. Nevertheless, more work is needed to understand the relationship between temporal clustering in acoustic onsets and linguistic units of analysis (e.g., see Kohler, 2008).

THE EMERGENCE OF MULTISCALE CLUSTERING OF PRELINGUISTIC VOCALIZATIONS

Additional work has focused on the emergence of multiscale clustering in infant speech development. A number of previous studies have documented infants' *perceptual* sensitivity to hierarchical structure in auditory stimuli including naturally occurring sounds (Gervain, Werker, & Geffen, 2014), speech (Hirsh-Pasek et al., 1987; Jusczyk et al., 1992), and even music (Fernald & Kuhl, 1987). One of the only published studies on infant *production* of hierarchically clustered vocalizations was conducted by Lynch, Oller, Steffens, and Buder (1995). In Lynch et al., infant vocalizations were recorded during free play, and utterances and syllables were later located by trained adult judges. The numeric timings of the utterances were printed out and given to untrained adult judges, who were tasked with bracketing the utterances that went together while listening to the recordings. The overall results from Lynch et al. were that prelinguistic infants' vocalizations span multiple scales of organization, from syllables, to utterances, to prelinguistic phrases. Therefore, even from very young ages, there is increasing evidence that speech perception and production is sensitive to multiscale patterns and multiscale patterns are present in prelinguistic vocal productions.

Methods and tools identifying hierarchical structure, multiscale clustering, and long-range correlations of human behavior have advanced considerably since Lynch et al. (1995), and our lab has focused efforts on a large-scale longitudinal corpus of infant-caregiver interactions (see also Johnson et al., 2013). From AF analysis of approximately 8,500 hr of recordings from 15 infant-adult dyads across the infants' first 2 years of life, we have found evidence for the multiscale clustering of prelinguistic vocal production (Abney, Warlaumont, Oller, Wallot, & Kello, 2015). Multiscale clustering is evident in all the recordings, including in the very youngest infant's recording, made at 11 days old. The timescales of this analysis spanned from ~10 s to ~1.5 hr, so reflect clustering at relatively larger timescales. In a longitudinal case study (Abney, Warlaumont, Haussman, Ross, & Wallot, 2014) of one of the infants, the largest change in the multiscale clustering of prelinguistic vocalizations directly preceded her parents' observations of canonical babbling. Canonical babbling is the production of syllables containing both consonant and vowel sounds and is considered an important prelinguistic vocal milestone (Oller, 2000; Patten et al., 2014). Overall, it is clear that infants have the capacity to produce nonrandom, multiscale clustered prelinguistic vocalizations at a very early age. This capacity may be important for language acquisition.

THE CONVERGENCE OF MULTISCALE CLUSTERING IN SPEECH ACOUSTICS

Our previous work established that the production of speech across a wide range of ages follows a pattern of multiscale clustering. A natural progression is to consider if the multiscale structures produced by two individuals during an interaction become correlated with each other. Previous studies established that, during dialogue, interlocutors match[1] properties of phonetic productions (Pardo, 2006, 2013), speech pauses (Capella & Planalp, 1981), syntactic structures (Bock, 1986), and lexical expressions of confidence (Fusaroli et al., 2012), and so on. Pickering and Garrod's (2004) interactive alignment model provides a framework for alignment within and across linguistic levels.

We extended AF analysis of multiscale clustering in speech to measure alignment of interlocutors across linguistic levels (Abney, Paxton, et al., 2014). Our analyses were inspired by recent theoretical work showing that maximal information exchange can occur between coupled complex systems, a phenomenon termed *complexity matching* (West, Geneston, & Grigolini, 2008;

[1]Alternatively known as alignment, convergence, coordination, entrainment, and so on (Louwerse, Dale, Bard, & Jeuniaux, 2012).

see also Aquino, Bologna, West, & Grigolini, 2011). In modeling the interaction of two complex networks, West and colleagues (e.g., Aquino et al., 2011) showed that maximal information transmission occurs when both systems exhibit the complex scaling patterns of 1/f-noise. If the two systems did not match scaling patterns, information transmission was not maximal and reduced as a function of the degree of matching. West and colleagues' formulation of complexity matching corresponds to a general hypothesis that organisms perceive and coordinate behaviors by exploiting invariant properties of the environment that include complex scaling patterns similar to what the AF analysis among other fractal and multifractal analyses measures. Therefore, the notion of complexity matching has been extended since West and colleagues' initial work to include the general hypothesis that organisms can synchronize and coordinate their behaviors with complex scaling patterns produced by their environments and other organisms. Cognitive scientists have begun using the concept of complexity matching to study human interaction across various experimental contexts and behavioral modalities (Abney, Paxton, et al., 2014; Coey, Washburn, & Richardson, 2014; Marmelat & Delignières, 2012; Paxton, Abney, Kello, & Dale, 2013).

In the same study discussed earlier (The Multiscale Clustering of Speech Production section), Abney, Paxton, et al. (2014) used AF analysis to measure the matching between multiscale speech structures produced by individuals in conversations. We found more matching for dyads in the affiliative conversation beyond what would be expected by chance based on a surrogate analysis. Above chance AF matching did not occur for dyads in the argumentative conversation. In other words, convergence of multiscale clustering of speech production only occurred for people in the affiliative conversation. Similar to how the multiscale structure of speech production was sensitive to conversational context, this work suggests that the amount of complexity matching is also context specific. Future work should further explore contextual differences in complexity matching.

For infant-caregiver interactions (Abney et al., 2015), there is also evidence for matching of multiscale clustering in prelinguistic vocalizations produced by infants with those produced by their caregivers. Notably, there is more complexity matching for speech-related vocalizations (speech, nonword babble, and singing) relative to nonspeechlike vocalizations (laughing, crying, burping, and coughing), suggesting that the matching process is sensitive to different types of vocal behaviors.

Recent empirical efforts applying the concept of complexity matching to human interaction has focused on the contexts where complexity matching occurs. However, one prediction of complexity matching (West et al., 2008) is that information transfer between two complex systems is maximal when the complexities of the systems are strongly coupled. Less work has focused on this prediction. In a reanalysis of speech signals similar to Abney, Paxton, et al. (2014)

from a joint perceptual decision-making task where dyads collaborated to make visual discrimination judgments (Bahrami et al., 2010), Fusaroli, Abney, Bahrami, Kello, and Tylén (2013) found that complexity matching correlated with higher performance on the task. These results suggest that stronger convergence of multiscale structure of vocal productions between interlocutors may have led to higher performance by facilitating information transfer. Additional work is necessary to test the prediction of maximal information transfer across strongly coupled complex systems as well as to relate mathematical notions of information transfer to more linguistic conceptions of semantic information transfer.

FUTURE DIRECTIONS

There is growing evidence for the multiscale clustering of speech production akin to other observations of similar complex patterns of human behavior including postural sway (Collins & De Luca, 1994; Delignières, Torre, & Bernard, 2011; Kelty-Stephen & Dixon, 2013), walking (Hausdorff et al., 2001; Marmelat, Delignières, Torre, Beek, & Daffertshofer, 2014), and exploratory movements of eyes (Aks, Zelinsky, & Sprott, 2002; Rhodes, Kello, & Kerster, 2014; Stephen & Mirman, 2010) and body (Palatinus, Kelty-Stephen, Kinsella-Shaw, Carello, & Turvey, 2014; Stephen, Arzamarski, & Michaels, 2010; Stephen & Hajnal, 2011). Specifically, the property of multiscale clustering of speech production adds to a list of power law functions in language behaviors such as frequencies of word usage (Zipf, 1949); temporal clustering of letters, words, and topics in text (Altmann, Cristadoro, & Esposti, 2012); and fluctuations in speech dynamics (Holden & Rajaraman, 2012; Kello, Anderson, Holden, & Van Orden, 2008).

Our analysis of speech signals has thus far considered a single power law scaling of clustered onsets of speech during human interaction. As discussed earlier, the AF is the point process analog to the interval-based analysis of DFA, which estimates the monofractal structure of a time series signal. The estimate of a monofractal structure in a system provides evidence for the "self-similarity" of behavior across different scales, which indexes how the structure (or variability) correlates with timescale. However, a system can also exhibit heterogeneous structure (or variability) across different scales while still displaying a strong correlation between degree of variability and analysis scales (Ihlen & Vereijken, 2010, 2013). Despite the emerging evidence for multifractal patterns in human behavior (Harrison, Kelty-Stephen, Vaz, & Michaels, 2014; Ihlen, 2014; Palatinus et al., 2014), less work has focused on the patterns of vocalizations that might entail multifractal structure (Gonzalez, Ling, & Violaro, 2012; Hasselman, 2015). Future work should focus on new methods for estimating the multiscale clustering of speech signals that exhibit nested structure and require a spectrum of power law exponents for description.

Furthermore, studying human interaction in terms of complexity matching would benefit from advancements in the measures of the complexity of the interacting systems, including multifractal formalisms (Ihlen & Vereijken, 2013). Although the first wave of research studying complexity matching in human interaction has demonstrated the phenomenon in various contexts using monofractal methods (Abney, Paxton, et al., 2014; Marmelat & Delignières, 2012; Paxton et al., 2013; but see Coey et al., 2014, for the application of recurrence quantification analysis), the complex coordination of human interaction might require more refined measures that take into consideration the heterogeneity of multiscale coordination.

Finally, the connections between complexity matching and strong anticipation[2] (Dubois, 2003; Stephen & Dixon, 2011; Stephen, Stepp, Dixon, & Turvey, 2008; Stepp & Turvey, 2010) have already led to interesting questions about the dependence on local or global coordination patterns (Fine, Likens, Amazeen, & Amazeen, 2015; Marmelat & Delignières, 2012; Torre, Varlet, & Marmelat, 2013). Future work should continue to focus on the question of local versus global coordination dependencies in addition to the differences of coordination patterns across modalities during an interaction.

CONCLUSION

The observation of multiscale clustering of speech production adds to the growing list of human behaviors exhibiting complex multiscale patterns. We provided a brief review of the current evidence of multiscale clustering of speech production across a variety of ages and interactional contexts. Additionally, we discussed the concept of complexity matching in human interaction and how this concept from statistical mechanics can be used to investigate the convergence of multiscale vocal productions across two people during an interaction. Human interaction is a complex coordination of human behavior spanning multiple timescales and modalities. This article described how one behavior—speech production—can entail a multiscale structure and how such structures from two people can converge during an interaction. We are optimistic that future researchers might

[2]The notion of strong anticipation originates from Dubois' (2003) distinction between weak and strong anticipation. In the context of an organism interacting with the environment, short-term anticipation based on the organism's internal model of the environment for prediction and local attunement of behavior is referred to as weak anticipation. The general concept of strong anticipation, in contrast, does not require an internal model. For strong anticipation, the coupling between organism and environment relies on global attunement to statistical properties of the environment, like the complex patterns exhibited by fractal scaling.

apply the ideas demonstrated here for speech production to other behaviors and coordination patterns during naturalistic interactions.

REFERENCES

Abney, D. H., Paxton, A., Dale, R., & Kello, C. T. (2014). Complexity matching in dyadic conversation. *Journal of Experimental Psychology. General*, *143*(6), 2304–2315.

Abney, D. H., Warlaumont, A. S., Haussman, A., Ross, J. M., & Wallot, S. (2014). Using nonlinear methods to quantify changes in infant limb movements and vocalizations. *Frontiers in Psychology*, *5*(771), 1–15. doi:10.3389/fpsyg.2014.00771

Abney, D. H., Warlaumont, A. S., Oller, D. K., Wallot, S., & Kello, C. T. (2015). *The multiscale clustering of infant vocalization bouts*. Manuscript submitted for publication.

Aks, D., Zelinsky, G., & Sprott, J. (2002). Memory across eye-movements: 1/f dynamic in visual search. *Nonlinear Dynamics, Psychology, and Life Sciences*, *6*(1), 1–25.

Allan, D. W. (1966). Statistics of atomic frequency standards. *Proceedings of the IEEE*, *54*, 221–230.

Altmann, E. G., Cristadoro, G., & Esposti, M. D. (2012). On the origin of long-range correlations in texts. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(29), 11582–11587. doi:10.1073/pnas.1117723109

Aquino, G., Bologna, M., West, B. J., & Grigolini, P. (2011). Transmission of information between complex systems: 1/f resonance. *Physical Review E*, *83*(5), 051130. doi:10.1103/PhysRevE.83.051130

Bahrami, B., Olsen, K., Latham, P. E., Roepstorff, A., Rees, G., & Frith, C. D. (2010). Optimally interacting minds. *Science*, *329*(5995), 1081–1085. doi:10.1126/science.1185718

Bak, P. (1996). *How nature works*. New York, NY: Springer New York. doi:10.1007/978-1-4757-5426-1

Bassingthwaighte, J. B., Liebovitch, L. S., & West, B. J. (2013). *Fractal physiology* (2nd ed.). New York, NY: Springer.

Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive Psychology*, *18*(3), 355–387. doi:10.1016/0010-0285(86)90004-6

Capella, J., & Planalp, S. (1981). Talk and silence sequences in informal conversations III: Interspeaker influence. *Human Communication Research*, *7*(2), 117–132. doi:10.1111/j.1468-2958.1981.tb00564.x

Clauset, A., Shalizi, C., & Newman, M. (2009). Power-law distributions in empirical data. *SIAM Review*, *51*, 661–703. doi:10.1137/ 070710111

Coey, C. A., Washburn, A., & Richardson, M. J. (2014). Recurrence quantification as an analysis of temporal coordination with complex signals. In N. Marwan, M. Riley, A. Giuliani, & C. L. Webber, Jr. (Eds.), Translational recurrences: From mathematical theory to real-world applications (pp. 173–186). Chaim, Switzerland: Springer.

Collins, J., & De Luca, C. (1994). Random walking during quiet standing. *Physical Review Letters*, *73*(5), 764–767. doi:10.1103/PhysRevLett.73.764

Cummins, F., & Port, R. (1998). Rhythmic constraints on stress timing in English. *Journal of Phonetics*, *26*(2), 145–171. doi:10.1006/jpho.1998.0070

Delignières, D., Torre, K., & Bernard, P.-L. (2011). Transition from persistent to anti-persistent correlations in postural sway indicates velocity-based control. *PLoS Computational Biology*, *7*(2), e1001089. doi:10.1371/journal.pcbi.1001089

Dubois, D. M. (2003). Mathematical foundations of discrete and functional systems with strong and weak anticipations. In M. V. Butz, O. Sigaud, & P. Gérard (Eds.), *Anticipatory behavior in adaptive learning systems* (pp. 110–132). Berlin and Heidelberg, Germany: Springer.

Fernald, A., & Kuhl, P. (1987). Acoustic determinants of infant preference for motherese speech. *Infant Behavior and Development*, *10*(3), 279–293. doi:10.1016/0163-6383(87)90017-8

Fine, J. M., Likens, A. D., Amazeen, E. L., & Amazeen, P. G. (2015). Emergent complexity matching in interpersonal coordination: Local dynamics and global variability. *Journal of Experimental Psychology: Human Perception and Performance*, *41*, 723–737. doi:10.1037/xhp0000046

Fusaroli, R., Abney, D. H., Bahrami, B., Kello, C. T., & Tylén, K. (2013, July). *Conversation, coupling, and complexity: matching scaling laws predicts performance in a joint decision task*. Poster presented at the 35th Annual Conference of the Cognitive Science Society, Berlin, Germany.

Fusaroli, R., Bahrami, B., Olsen, K., Roepstorff, A., Rees, G., Frith, C., & Tylén, K. (2012). Coming to terms: Quantifying the benefits of linguistic coordination. *Psychological Science*, *23*(8), 931–939. doi:10.1177/0956797612436816

Gervain, J., Werker, J. F., & Geffen, M. N. (2014). Category-specific processing of scale-invariant sounds in infancy. *PloS One*, *9*(5), e96278. doi:10.1371/journal.pone.0096278

Gonzalez, D. C., Ling, L. L., & Violaro, F. (2012). Analysis of the multifractal nature of speech signals. In L. Alvarez, M. Mejail, L. Gomez, & J. Jacobo (Eds.), *Progress in pattern recognition, image analysis, computer vision, and applications* (Vol. 7441, pp. 740–748). Berlin and Heidelberg, Germany: Springer. doi:10.1007/978-3-642-33275-3

Harrison, H. S., Kelty-Stephen, D. G., Vaz, D. V., & Michaels, C. F. (2014). Multiplicative-cascade dynamics in pole balancing. *Physical Review E*, *89*(6), 060903. doi:10.1103/PhysRevE.89.060903

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

Hausdorff, J. M., Ashkenazy, Y., Peng, C.-K., Ivanov, P. C., Stanley, H. E., & Goldberger, A. L. (2001). When human walking becomes random walking: Fractal analysis and modeling of gait rhythm fluctuations. *Physica A: Statistical Mechanics and Its Applications*, *302*(1–4), 138–147. doi:10.1016/S0378-4371(01)00460-5

Hirsh-Pasek, K., Kemler Nelson, D. G., Jusczyk, P. W., Cassidy, K. W., Druss, B., & Kennedy, L. (1987). Clauses are perceptual units for young infants. *Cognition*, *26*(3), 269–286. doi:10.1016/S0010-0277(87)80002-1

Holden, J. G., & Rajaraman, S. (2012). The self-organization of a spoken word. *Frontiers in Psychology*, doi:10.3389/fpsyg.2012.00209

Ihlen, E. A. F. (2014). Age-related changes in inter-joint coordination during walking. *Journal of Applied Physiology*, *117*(2), 189–198. doi:10.1152/japplphysiol.00212.2014

Ihlen, E. A. F., & Vereijken, B. (2010). Interaction-dominant dynamics in human cognition: Beyond 1/f(alpha) fluctuation. *Journal of Experimental Psychology: General*, *139*(3), 436–463. doi:10.1037/a0019098

Ihlen, E. A. F., & Vereijken, B. (2013). Multifractal formalisms of human behavior. *Human Movement Science*, *32*(4), 633–651. doi:10.1016/j.humov.2013.01.008

Johnson, N. F., Medina, P., Zhao, G., Messinger, D. S., Horgan, J., Gill, P., … Zarama, R. (2013). Simple mathematical law benchmarks human confrontations. *Scientific Reports*, *3*, 3463. doi:10.1038/srep03463

Jusczyk, P. W., Hirsh-Pasek, K., Kemler Nelson, D. G., Kennedy, L. J., Woodward, A., & Piwoz, J. (1992). Perception of acoustic correlates of major phrasal units by young infants. *Cognitive Psychology*, *24*(2), 252–293. doi:10.1016/0010-0285(92)90009-Q

Kello, C. T., Anderson, G. G., Holden, J. G., & Van Orden, G. C. (2008). The pervasiveness of 1/f scaling in speech reflects the metastable basis of cognition. *Cognitive Science*, *32*(7), 1217–1231. doi:10.1080/03640210801944898

Kello, C. T., Beltz, B. C., Holden, J. G., & Van Orden, G. C. (2007). The emergent coordination of cognitive function. *Journal of Experimental Psychology: General*, *136*(4), 551–568.

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

Kelty-Stephen, D. G., & Dixon, J. A. (2013). Temporal correlations in postural sway moderate effects of stochastic resonance on postural stability. *Human Movement Science*, *32*, 91–105.

Kohler, K. J. (2008). The perception of prominence patterns. *Phonetica*, *65*(4), 257–269. doi:10.1159/000192795

Louwerse, M. M., Dale, R., Bard, E. G., & Jeuniaux, P. (2012). Behavior matching in multimodal communication is synchronized. *Cognitive Science*, *36*(8), 1404–1426. doi:10.1111/j.1551-6709.2012.01269.x

Lowen, S. B., & Teich, M. C. (1996). The periodogram and Allan variance reveal fractal exponents greater than unity in auditory-nerve spike trains. *The Journal of the Acoustical Society of America*, *99*(6), 3585–3591.

Lynch, M. P., Oller, D. K., Steffens, M. L., & Buder, E. H. (1995). Phrasing in prelinguistic vocalizations. *Developmental Psychobiology*, *28*(1), 3–25. doi:10.1002/dev.420280103

Mandelbrot, B. B. (1983). *The fractal geometry of nature*. San Francisco, CA: Freeman.

Marmelat, V., & Delignières, D. (2012). Strong anticipation: Complexity matching in interpersonal coordination. *Experimental Brain Research*, *222*(1–2), 137–148. doi:10.1007/s00221-012-3202-9

Marmelat, V., Delignières, D., Torre, K., Beek, P. J., & Daffertshofer, A. (2014). "Human paced" walking: Followers adopt stride time dynamics of leaders. *Neuroscience Letters*, *564*, 67–71. doi:10.1016/j.neulet.2014.02.010

Oller, D. K. (2000). *The emergence of the speech capacity*. New York, NY: Psychology Press.

Palatinus, Z., Kelty-Stephen, D. G., Kinsella-Shaw, J., Carello, C., & Turvey, M. T. (2014). Haptic perceptual intent in quiet standing affects multifractal scaling of postural fluctuations. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(5), 1808–1818.

Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America*, *119*(4), 2382, doi:10.1121/1.2178720

Pardo, J. S. (2013). Measuring phonetic convergence in speech production. *Frontiers in Psychology*, doi:10.3389/fpsyg.2013.00559

Patten, E., Belardi, K., Baranek, G. T., Watson, L. R., Labban, J. D., & Oller, D. K. (2014). Vocal patterns in infants with autism spectrum disorder: Canonical babbling status and vocalization frequency. *Journal of Autism and Developmental Disorders*, *44*, 2413–2428. doi:10.1007/s10803-014-2047-4

Paxton, A., Abney, D. H., Kello, C. T., & Dale, R. (2013). Network analysis of multimodal, multiscale coordination in dyadic problem solving. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th Annual Meeting of the Cognitive Science Society* (pp. 2735–2740). Austin, TX: Cognitive Science Society.

Paxton, A., & Dale, R. (2013). Argument disrupts interpersonal synchrony. *The Quarterly Journal of Experimental Psychology*, *66*, 2092–2102. doi:10.1080/17470218.2013.853089

Peng, C. K., Havlin, S., Stanley, H. E., & Goldberger, A. L. (1995). Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series. *Chaos*, *5*(1), 82–87. doi:10.1063/1.166141

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*(2), 169–190.

Rhodes, T., Kello, C. T., & Kerster, B. (2014). Intrinsic and extrinsic contributions to heavy tails in visual foraging. *Visual Cognition*, *22*(6), 809–842. doi:10.1080/13506285.2014.918070

Stephen, D. G., Arzamarski, R., & Michaels, C. F. (2010). The role of fractality in perceptual learning: Exploration in dynamic touch. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(5), 1161–1173. doi:10.1037/a0019219

Stephen, D. G., & Dixon, J. A. (2011). Strong anticipation: Multifractal cascade dynamics modulate scaling in synchronization behaviors. *Chaos, Solitons & Fractals*, *44*(1–3), 160–168. doi:10.1016/j.chaos.2011.01.005

Stephen, D. G., & Hajnal, A. (2011). Transfer of calibration between hand and foot: Functional equivalence and fractal fluctuations. *Attention, Perception & Psychophysics*, *73*(5), 1302–1328. doi:10.3758/s13414-011-0142-6

Stephen, D. G., & Mirman, D. (2010). Interactions dominate the dynamics of visual cognition. *115*(1), 154–165. doi:1 10.1016/j.cognition.2009.12.010

Stephen, D. G., Stepp, N., Dixon, J. A., & Turvey, M. T. (2008). Strong anticipation: Sensitivity to long-range correlations in synchronization behavior. *Physica A: Statistical Mechanics and Its Applications*, *387*(21), 5271–5278. doi:10.1016/j.physa.2008.05.015

Stepp, N., & Turvey, M. T. (2010). On strong anticipation. *Cognitive Systems Research*, *11*(2), 148–164. doi:10.1016/j.cogsys.2009.03.003

Thurner, S., Lowen, S. B., Feurstein, M. C., Heneghan, C., Feichtinger, H. G., & Teich, M. C. (1997). Analysis, synthesis, and estimation of fractal-rate stochastic point processes. *Fractals*, *5*, 565–595. doi:10.1142/S0218348X97000462

Tilsen, S. (2009). Multitimescale dynamical interactions between speech rhythm and gesture. *Cognitive Science*, *33*(5), 839–879. doi:10.1111/j.1551-6709.2009.01037.x

Torre, K., Varlet, M., & Marmelat, V. (2013). Predicting the biological variability of environmental rhythms: Weak or strong anticipation for sensorimotor synchronization? *Brain and Cognition*, *83*(3), 342–350. doi:10.1016/j.bandc.2013.10.002

Van Orden, G. C., Holden, J. G., & Turvey, M. T. (2003). Self-organization of cognitive performance. *Journal of Experimental Psychology: General*, *132*(2), 331–350.

Viswanathan, G. M., Peng, C. K., Stanley, H. E., & Goldberger, A. L. (1997). Deviations from uniform power law scaling in nonstationary time series. *Physical Review E: Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, *55*(1), 845–849.

West, B. J. B., Geneston, E. L., & Grigolini, P. (2008). Maximizing information exchange between complex networks. *Physics Reports*, *468*(1–3), 1–99. doi:10.1016/j.physrep.2008.06.003

Zipf, G. K. (1949). *Human behavior and the principle of least effort*. Reading, MA: Addison-Wesley.

APPENDIX

The multiscale clustering of vocalizations is estimated using Allan Factor (AF) analysis. Each time series of acoustic onsets is segmented into M adjacent and nonoverlapping windows of size T, then the number of events N_j is counted within each window indexed by $j = 1$ to M. The differences in counts between adjacent windows of a given size T is computed as $d(T) = N_{j+1}(T) - N_j(T)$. The AF variance $\mathrm{A}(T)$ for a given timescale, T, is the mean value of the squared differences, normalized by mean counts of events per window (i.e., closely related to coefficient of variation),

$$A(T) = \frac{\langle d(T)^2 \rangle}{2\langle N(T) \rangle}.$$

Poisson processes (i.e., random, independent events with exponentially distributed interevent intervals) yield $\mathrm{A}(T) \sim 1$ for all T. In contrast, power law clustering yields $\mathrm{A}(T) > 1$, specifically with $\mathrm{A}(T) \sim (T/T_1)^{\alpha}$, where T_1 is the smallest timescale considered and α the exponent of the scaling relation (Thurner et al., 1997).

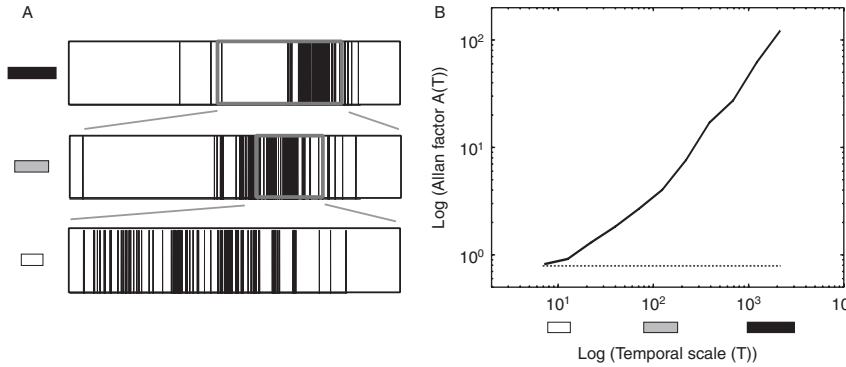

FIGURE A1 Schematic depiction of procedure of Allan Factor (AF) analysis. (A) Vocalization events are counted within each timescale window. Each vertical line is an acoustic onset. The black, grey, and white rectangles indicate long (~60 min), medium (~30 min), and short (~7 min) timescales, respectively. Notice at each of the three timescales, there are clusters of onsets. (B) The estimates of multiscale clustering of vocalizations. AF variance is derived from computing the squared difference of onset frequencies between adjacent time windows for the three timescales. The slope of the log(AF) versus log(T) curve estimates the scaling of AF variance across scales. The curve with dotted line indicates a slope ~0, which is evidence for a random (Poisson process) vocalization event series. The other slope is closer to 1, indicating multiscale clustering.

Multiscale clustering is therefore indicated when $A(T) \sim T^{\alpha}$, where $\alpha > 0$. This is a power law with exponent α, which provides a metric for the degree to which events are clustered across timescales. α corresponds to the slope of the plot in panel B of Figure A1, which plots coefficient of variation versus timescale on a log-log plot. The further α is from 0 and the closer it is to 1, the more structured we say the clustering of vocalizations is across scales (see Figure A1).