Title
The Complexity Matching hypothesis for human communication

Author

Abney, Drew Hamilton

Publication Date

2016

UNIVERSITY OF CALIFORNIA, MERCED

The Complexity Matching hypothesis for human communication

A dissertation submitted in partial satisfaction of the requirements
for the degree Doctor of Philosophy

in

Cognitive and Information Sciences

by

Drew H. Abney

Committee in charge:

Professor Christopher T. Kello, Chair
Professor Ramesh Balasubramaniam
Professor Anne S. Warlaumont

2016

The dissertation of Drew H. Abney is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Professor Christopher T. Kello, Chair

Professor Ramesh Balasubramaniam

Professor Anne S. Warlaumont

University of California, Merced

2016

This dissertation is dedicated to my family:

to my wife, Rebecca, who has given me endless love and support;

and to my parents who provided the countless opportunities for me to grow.

My achievements would be impossible without their encouragement and love.

Contents

List of Figures viii

List of Tables xi

Acknowledgements xii

Curriculum Vita xiii

Abstract xxv

1 Introduction 1
- 1.1 The production and convergence of hierarchical patterns of communicative behavior . . . 3
- 1.2 The Complexity Matching hypothesis for human communication . . . 4
- 1.3 The present work . . . 6

2 Complexity matching in dyadic conversation 8
- 2.1 Preface . . . 8
- 2.2 Introduction . . . 8
- 2.3 Power law clustering in conversational speech . . . 11
- 2.4 Complexity matching in speech signal clustering . . . 12
- 2.5 Current Study . . . 17
- 2.6 Methods . . . 18
  - 2.6.1 Participants . . . 18
  - 2.6.2 Procedure . . . 18
  - 2.6.3 Apparatus, data collection, and data preparation . . . 19
- 2.7 Inter-event intervals . . . 20
- 2.8 Temporal clustering in acoustic onsets . . . 22
- 2.9 Formal description of AF analysis . . . 22
- 2.10 Results of AF analysis . . . 24
- 2.11 Complexity matching . . . 24
- 2.12 Behavioral matching . . . 25

2.13 General Discussion 28
2.13.1 Complexity matching and theories of conversation, coordination, and development 29
2.14 Conclusion 32

3 Multimodal complexity matching and information transmission in a dyadic problem-solving task 33
3.1 Preface 33
3.2 Introduction 33
3.3 Method 35
3.3.1 Participants 35
3.3.2 Materials and procedure 36
3.3.3 Vocalization and Movement analyses 36
3.4 Results 40
3.5 Discussion and Conclusions 40

4 Multiple coordination patterns in infant and adult vocalizations 44
4.1 Preface 44
4.2 Introduction 45
4.2.1 Goals of the current study 47
4.3 Method 48
4.3.1 Participants 48
4.3.2 Data Collection 48
4.3.3 Analyses 50
4.4 Results 52
4.4.1 Volubility and Hierarchical Clustering Across Vocalization Types 52
4.4.2 Do coincidence-based, rate-based, and cluster-based coordination patterns vary depending on the type of vocalization produced by the infant? 56
4.4.3 Are adults or infants primarily driving these vocal coordination patterns, and does this change with age? 59
4.4.4 Do the different coordination measures have unique developmental trends? 62
4.5 Discussion 64
4.5.1 Hierarchical vocalization patterns and volubility . . . 65
4.5.2 Vocal coordination patterns vary by vocalization type and provide unique information based on level of description 66
4.5.3 Different coordination patterns provide unique information about the dynamics of vocal interaction . . . 67
4.5.4 Coordination patterns and infant age 68

4.5.5 Future directions 69
4.6 Conclusion 70

5 Discussion **71**
5.1 Introduction 71
5.2 Production and convergence of hierarchical structure 71
5.3 Quantification of multiscale clustering of communicative behaviors 73
5.4 Information and information transmission 74
5.5 Development of hierarchical communicative structure 75
5.6 Conclusion 76

References **77**

List of Figures

2.1 (Left) Pickering and Garrod's (2004, reprinted with permission) schematic representation of the stages of comprehension and production processes according to the interactive alignment model. (Right) An illustration of the nesting of different scales of linguistic representations, using four levels of the six from the interactive alignment model: phonetic, lexical, semantic, and situation model. 12
2.2 (Left) An example conversational speech signal, shown at three different temporal scales. (A) The longest scale roughly corresponds with conversational turns. The phonetic, lexical, semantic, and situation model labels approximate the time scales of these units on the speech signal. (B) The middle scale roughly corresponds with e.g. thinking pauses and phrase boundaries. (C) The shortest scale roughly corresponds with word, syllable, and phoneme boundaries. Vertical lines show acoustic onsets relative to a threshold chosen by visual inspection. 14
2.3 (Left) Examples of synchronization and behavior matching with toy metronome systems. (A) Illustration of two metronomes interacting along a sliding platform, as a simple model of synchronization and a form of behavioral matching. (B) Illustration of interactions between multiple metronomes with differing frequencies, to aid the intuition of complexity matching. 16
2.4 (Left) IEI probability density functions for individual interlocutors in individual conversations, plotted in logarithmic coordinates using logarithmic binning. Dashed line shows idealized slope of -2 (per West et al., 2008). 21
2.5 (Left) Mean AF functions for argumentative vs. affiliative conversation types, with standard error bars. 23

2.6 (Left) Mean summed AF difference functions plotted for the two conversation types, separately for original pairings versus randomized controls, with standard error bars. 26

3.1 (Top) Example standardized movement difference series. Horizontal line represents event threshold. (Bottom) Example event series. 37
3.2 Mean Allan Factor functions for the Vocalization, Movement, and Vocalization/Movement time series. Error bars represent standard error. 39
3.3 Complexity matching predicting tower height (cm). Individual dots correspond to each trial. Lines represent linear fit. 41

4.1 Schematic depiction of procedure of AF analysis at three timescales (∼7 minutes, ∼30 minutes, ∼60 minutes). (A-C) Vocalization events are counted within each timescale window. Each vertical line is an acoustic onset for one of the three vocalization types: (A) Infant speech-related, (B) Infant non-speech-related, and (C) Adult. The black, grey, and white rectangles indicate long (∼60 minutes), medium (∼30 minutes), and short timescales (∼7 minutes), respectively. Notice at each of the three timescales, there are clusters of onsets. AF variance is derived from computing the normalized squared difference of onset frequencies between adjacent time windows for the three timescales. AF variance is a measure of the departure from an equidistributed distribution of acoustic onsets. (D) The estimates of hierarchical clustering of vocalization types. The slope, α, of the $\log(AF)$ vs. $\log(T)$ curve estimates the scaling of AF variance across scales. The dotted line indicates a slope of 0 which is evidence for a random (Poisson process) vocalization event series. The other three curves have slopes closer to 1, indicating hierarchical clustering. 53
4.2 (A) Mean AF functions for adult and infant vocalizations, with standard error bars. (B) Scatterplot of each recording's A(T) values. 54
4.3 Diagonal cross-recurrence profile (DCRP) averaged across all vocalization types. (Left) Average DCRPs are before normalization. (Right) Average DCRPs normalized for shuffled DCRPs. 58

4.4 (Top row) Cluster-based vocal coordination results for Adult and (left to right) Infant-combined, Infant-speech-related, and Infant-non-speech-related. (Bottom row) Rate-based vocal coordination results. All variables are standardized. Each circle represents an individual recording. 60
4.5 (Top row) Difference Score (DS) results for infant age and (left to right) Infant-combined hierarchical clustering estimates, speech-related hierarchical clustering estimates, and non-speech-related hierarchical clustering estimates. (Bottom row) DS results for infant age and (left to right) infant-combined volubility, speech-related volubility, and non-speech-related volubility. Note. AF and Volubility DS axes have different ranges. 63

List of Tables

4.1 Results of first order correlations and residual analyses predicting infant age. 56
4.2 Results of first order correlations (r) and residual correlations (r_{residual}) predicting matching of infant vocalization properties with adult volubility and adult AF slope estimates. 59
4.3 Results of first order correlations (r) and residual correlations (r_{residual}) of coordination patterns and infant age. . . . 64

Acknowledgements

I would like to thank my committee members – Chris Kello, Ramesh Balasubramaniam, and Anne Warlaumont – for their support and mentorship during my graduate studies at UC Merced. Each member provided unique lessons, encouragement, and critical thoughts that have undoubtedly made me a better scientist.

Thanks to my main advisor and mentor, Chris Kello, for always providing a supportive academic environment to grow and expand my curiosity. From the beginning of my graduate career at UC Merced, Chris has provided endless encouragement, allowing me to explore new ideas, which ultimately culminated into the body of work in this dissertation.

Thanks to Ramesh Balasubramaniam for the wonderful and important conversations about science, politics, and coffee. From our first interactions, Ramesh treated me as a colleague and provided invaluable insights for a young scientist. Thanks to Anne Warlaumont for providing a model of a dedicated, young professor. Anne started her assistant professorship the same year I started my graduate studies at Merced. I have had the awesome opportunity to observe what it takes to become a rising star in our field.

I would also like to acknowledge previous mentors and colleagues who were influential in my graduate studies at Illinois State University. Chris Merrill and Josh Brown provided me with the initial launchpad into graduate research. Jeff Wagman and Dawn McBride have given me endless mentorship, encouragement, and opportunities. J. Scott Jordan has shown me how to be groovy and think deeply about the connections between science and experience.

I would also like to thank the UC Merced Cognitive and Information Sciences group for providing an exciting intellectual community for me to test and grow my ideas. From SSM to J&R's and from Coffee Bandits to the 17th Street Pub, there was always someone in our program willing to talk and debate ideas.

Finally, I would like to acknowledge the love and support from my wife, Rebecca, and my family. I met Rebecca during our first week at UC Merced. I could not imagine this journey without her.

Curriculum Vita

Drew H. Abney

drewabney@gmail.com
http://drewabney.weebly.com

Education

2012–2016	Ph.D in Cognitive and Information Sciences University of California, Merced
2010–2012	M.S. in Experimental Psychology (Dual Degrees: Cognitive/Behavioral Sciences and Quantitative Psychology) Illinois State University
2009–2010	M.S. in Science, Technology, Education, and Mathematics Education and Leadership Illinois State University
2004–2008	B.S. in Technology Education Illinois State University

---

Peer-Reviewed Journal Articles

Ross, J.M., Warlaumont, A.S., **Abney, D.H.**, Rigoli, L.M. & Balasubramaniam, R. (2016). Influence of Musical Groove on Postural Sway. *Journal of Experimental Psychology: Human Perception and Performance.*

Abney, D.H., & Wagman, J.B. (2015). Direct learning in auditory perception: An information-space analysis of perception of object length by sound. *Ecological Psychology.*

Abney, D.H., Warlaumont, A.S., & Kello, C.T. (2015). Production and convergence of multiscale clustering of speech. *Ecological Psychology.*

Abney, D.H., Paxton, A., Dale, R., & Kello, C. (2015). Movement dynamics reflect a functional role for weak coupling and role structure in dyadic problem solving. *Cognitive Processing.*

Abney, D.H., McBride, D.M., Conte, A. & Vinson, D.W. (2014). Response dynamics in prospective memory. *Psychonomic Bulletin & Review.*

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

Abney, D.H., Warlaumont, A.S., Haussmann, A., Ross, J., Wallot, S. (2014). Using nonlinear methods to quantify infant motor and vocal developmen. *Frontiers in Developmental Psychology, 143*(6), 2304-2315.

Vinson, D.W., **Abney, D.H.**, Warlaumont, A.S., Dale, R., & Matlock, T. (2014). The influence of linguistic and social information on visual memory. *Frontiers in Cognitive Science.*

Abney, D.H., Dale, R., Yoshimi, J., Fusaroli, R., Kello, C.T., & Tylen, K. (2014). Joint perceptual decision-making: A case study in explanatory pluralism. *Frontiers in Theoretical and Philosophical Psychology.*

Abney, D.H., Wagman, J.B., & Schneider, J. (2014). Changing grasp position on a wielded object provides self-training for perception of length. *Attention, Perception, & Psychophysics.*

Wagman, J.B. & **Abney, D.H.**. (2013). Is calibration of the perception of length modality-independent? *Attention, Perception, & Psychophysics.*

Abney, D.H., McBride, D.M., & Petrella, S.N. (2013). Interactive effects in transfer-appropriate processing for event-based prospective mem-

ory. *Memory & Cognition.*

McBride, D.M. & **Abney, D.H.**. (2012). A comparison of transfer-appropriate processing and multi-process frameworks for prospective memory performance. *Experimental Psychology.*

Wagman, J.B. & **Abney, D.H.**. (2012). Transfer of recalibration from audition to touch: Modality independence as a special case of anatomical independence. *Journal of Experimental Psychology: Human Perception and Performance. 38*(3), 589-602.

McBride, D.M., Beckner, J. & **Abney, D.H.**. (2011). Effect of placement of prospective memory cues in an ongoing task on prospective memory task performance. *Memory & Cognition. 39*, 1222-1231.

Manuscripts Under Review/Revision

Abney, D.H., Warlaumont, A.S., Oller, D.K., Wallot, S. & Kello, C.T. (under revision). The multiscale clustering of infant vocalization bouts.

Abney, D.H., Kello, C.T., & Balasubramaniam, R. (under revision). Introduction and application of the multiscale coefficient of variation analysis.

Abney, D.H., Gann, T.M., Huette, S., Matlock, T. (under revision). The Language of Uncertainty and Political Ideology in Climate Change.

Edited Books

Weast-Knapp, J., Malone, M., & **Abney, D.H.** (Eds.)(2015). *Studies in Perception and Action XIII. Proceedings from the Eighteenth International Conference on Perception and Action, Minneapolis, MN.*

Commentary

Vinson, D.W., **Abney, D.H.**, Amso, D., Anderson, M.L., Chemero, T., Cutting, J.E., Dale, R., Richardson, D., Friston, K., Gallagher, S., Jordan, J.S., Mudrik, L., Ondobaka, S., Shams, L., Shiffrar, M., Spivey, M.J. (2015). Perception, as you make it. Commentary of C. Firestone & B. Scholl's "Cognition does not affect perception: Evaluating the evidence for 'top-down' effects". *Behavioral and Brain Sciences*

Refereed Conference Proceedings

Bunce, J.P., **Abney, D. H.**, Gordon, C.L., Spivey, M.J., & Scott, R.M. (forthcoming). Using Motor Dynamics to Explore Real-time Competition in Cross-situational Word Learning: Evidence From Two Novel Paradigms. In J. Trueswell, A. Papafragou, D. Grodner, & D. Mirman (Eds.), *Proceedings of the 36th Annual Meeting of the Cognitive Science Society.* Austin, TX: Cognitive Science Society.

Abney, D. H. & Thomas, B. (2015). Exploration in information space during affordance perception. In Weast-Knapp, J., Malone, M., & Abney, D.H. (Eds.), *Studies in Perception and Action XIII: Proceedings from the Eighteenth International Conference on Perception and Action.*

Schloesser, D.S., Wagman, J.B. & **Abney, D. H.** (2015). Flip this rod! Changing grasp position can recalibrate perception of length by dynamic touch. In Weast-Knapp, J., Malone, M., & Abney, D.H. (Eds.), *Studies in Perception and Action XIII: Proceedings from the Eighteenth International Conference on Perception and Action.*

Paxton, A., **Abney, D, H.**, Kello, C. T., & Dale, R. (2014). Network analysis of multimodal, multiscale coordination in dyadic problem solving. In P. M. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th Annual Meeting of the Cognitive Science Society.* Austin, TX: Cognitive Science Society.

Abney, D. H., Paxton, A., Kello, C., & Dale, R. (2013). Complexity matching in dyadic interactions. In P. Passos, J. Barrieros, R. Cordovil, D. Araújo, & F. Melo (Eds.), *Studies in Perception and Action XII: Proceedings from the Seventeenth International Conference on Perception and Action.*

Conference Presentations

Abney, D.H., Warlaumont, A.S., Oller, D.K., Wallot, S., & Kello, C.T. (May, 2016). *The multiscale clustering of infant vocalization bouts.* Presentation at the International Conference on Infant Studies. New Orleans, LA.

Perlman, M. & **Abney, D.H.**, (March, 2016). *Multimodal analysis of the learned vocal repertoire of a human-fostered gorilla.* Presentation at the 2016 EvoLang conference. New Orleans, LA.

Abney, D.H. & Balasubramaniam, R. (2015, November). *Introduction and applications to the multiscale coefficient of variation analysis (MSCV).* Paper presented at the 45th Annual Meeting of the Society for Computers in Psychology. Chicago, IL.

Abney, D.H. & Kello, C.T. (2014, August). *Complexity matching in dyadic conversation.* Presentation at the 10th Annual Guy Van Orden UConn Workshop on Cognition and Dynamics. University of Connecticut, Storrs, CT.

Abney, D.H., Warlaumont, A.S., Oller, D.K., Wallot, S., & Kello, C.T. (2015). *The multiscale clustering of infant vocalization bouts.* Presentation at the Society for Complex Systems in Cognitive Science. Long Beach, CA.

Abney, D.H., Warlaumont, A.S., Oller, D.K., Wallot, S., & Kello, C.T. (2015). *The multiscale clustering of infant vocalization bouts.* Presentation at the International Society for Perception and Action. Minneapolis, MN.

Abney, D.H. (2014). *Complexity matching across conversational settings.* Presentation at the Finding Common Ground: Social, Ecological, and Cognitive Perspectives on Language Use. University of Connecticut, Storrs, CT.

Abney, D.H. & Kello, C.T. (2014, November). *Measuring multiscale temporal clustering in speech signals.* Paper presented at the 44th Annual Meeting of the Society for Computers in Psychology. Long Beach, CA.

Abney, D.H., Kerster, B.E. & Kello, C.T. (2014). *Multiscale spatial distributions of eye fixations decouple over time.* Presentation at the International Society of Ecological Psychology.

Abney, D.H. & Wagman, J.B. (2014). *An information-space analysis of perception of object length by sound.* Presentation at the International Society of Ecological Psychology.

Abney, D.H. & Kello, C.T. (2013). *Dynamics of social cognition.* Presentation and workshop given by C. Kello at the National University of Galway, Complex Systems Research Centre, Dynamical Analyses of Social Behaviour Workshop.

Abney, D.H., Paxton, A., Kello, C.T. & Dale, R. (2013). *Complexity matching in dyadic interaction.* Paper presented at the 17^{th} International Conference of Perception and Action. Lisbon, Portugal.

Vinson, D.W. & **Abney, D.H.** (2013). *The influence of linguistic and social information on visual memory.* Presentation at the 11^{th} International Cognitive Linguistics Conference. Edmonton, Alberta, Canada.

Abney, D.H. & Wagman, J.B. (2012). *Intrinsic feedback in recalibration of length perception.* Presentation at the 2012 Illinoisy Data Conference, Edwardsville, IL.

Abney, D.H. & Wagman, J.B. (2011). *Transfer of recalibration across modalities: Feedback on audition improves haptic perception.* Presentation at the 2011 Illinoisy Data Conference, Carbondale, IL.

Brown, J.W. & **Abney, D.H.** (2010). *The biotechnology of technology education: A descriptive investigation of the conceptual foundations of biotechnology necessary for technology education teachers.* . Presentation at the annual conference of the International Technology and Engineering Education Association, Charlotte, NC.

Refereed Posters

Schloesser, D.S., Bai, J., **Abney, D.H.** & Jordan, J.S. (2015). *Comparing performance and coordination dynamics of dyads and individuals in a complex control task.* Poster presented at the annual conference of the Psychonomics Society, Chicago, IL.

Conte, A., **Abney, D.H.** & McBride, D. (2015). *Response Dynamics in Time- and Event-based Prospective Memory.* Poster presented at the annual conference of the Psychonomics Society, Chicago, IL.

Abney, D.H., Warlaumont, A.S., Oller, D.K., Wallot, S., & Kello, C.T. (2015). *The multiscale clustering of infant vocalization bouts.* Poster presented at the 37^{th} Annual Meeting of the Cognitive Science Society, Long Beach, CA.

Schloesser, D.S., Bai, J., **Abney, D.H.** & Jordan, J.S. (2015). *Comparing performance and coordination dynamics of dyads and individuals in a complex control task.* Poster presented at the 37^{th} Annual Meeting of the Cognitive Science Society, Long Beach, CA.

Abney, D.H. & Thomas, B. (2015). *Exploration in information space during affordance perception.* Poster presented at the 18^{th} International Conference on Perception and Action

Schloesser, D.S., Wagman, J.B. & **Abney, D. H.** (2015). Flip this rod! Changing grasp position can recalibrate perception of length by dynamic touch. Poster presented at the 18^{th} International Conference on Perception and Action

Abney, D.H., Gann, T.M., Huette,S. & Matlock,T. (2015). *The language of uncertainty and political ideology in climate change.* Poster presented at the 13^{th} Biennial Conference on Communication and Environment, University of Colorado, Boulder, CO.

Abney, D.H. & Wagman, J.B. (2014). *Direct learning in auditory perception: An information-space analysis of auditory perception of object length.* Poster presented at the annual conference of the Psychonomics Society, Long Beach, CA.

Paxton, A., **Abney, D.H.**, Kello, C.T. & Dale, R. (2014). *Network analysis of multimodal, multiscale coordination in dyadic problem solving.* Poster presented at the 36^{th} Annual Meeting of the Cognitive Science Society.

Cho, K. **Abney, D.H.**, Brotman, R. & Feldman, L.B. (2014). *Articulatory and phonological codes interact in memory.* Poster presented at the 36^{th} Annual Meeting of the Cognitive Science Society.

Abney, D.H., Paxton, A., Kello, C.T. & Dale, R. (2014). *Multimodal and multiscale interpersonal interaction in a joint problem solving task.* Poster presented at the 26^{th} Annual Meeting of the American Psychological Society, San Francisco, CA.

Abney, D.H., Paxton, A., Kello, C.T. & Dale, R. (2013). *Complexity matching in dyadic interaction.* Poster presented at the 35^{th} Annual Meeting of the Cognitive Science Society.

Fusaroli, R., **Abney, D.H.**, Bahrami, B., Kello, C.T. & Tylen, K. (2013). *Performance in a joint decision task is predicted by the degree of complexity matching of interlocutors' speech events.* Poster presented at the 35^{th} Annual Meeting of the Cognitive Science Society.

Vinson, D.W., **Abney, D.H.**, Dale, R. & Matlock, T. (2013). *Visuospatial memory's sensitivity to language and image orientation.* Poster presented at the 35^{th} Annual Meeting of the Cognitive Science Society.

Wagman, J.B. & **Abney, D.H.** (2012). *Is calibration of perception modality independent?.* Poster presented at the annual conference of the Psychonomics Society, Minneapolis, MN.

Abney, D.H. & McBride, D. (2012). *A test of the graded match in processing view of event-based prospective memory.* Poster presented at the annual Illinois State University Graduate Research Symposium.

Abney, D.H. & Wagman, J.B. (2012). *Intrinsic feedback in recalibration of length perception.* Poster presented at the annual Illinois State University Graduate Research Symposium.

Abney, D.H. & Wagman, J.B. (2011). *The role of intrinsic, self- feedback in recalibration of length perception in dynamic touch.* Poster presented at the annual conference of the Psychonomics Society, Seattle, WA.

Abney, D.H. & Wagman, J.B. (2010). *Transfer of recalibration across modalities: Feedback on audition improves haptic perception.* Poster presented at the annual conference of the Psychonomics Society, Chicago, IL.

Abney, D.H. & Wagman, J.B. (2010). *Transfer of recalibration across modalities: Feedback on audition improves haptic perception.* Poster presented at the annual Illinois State University Graduate Research Symposium.

Abney, D.H., Cantin, R.H., Merrill, C. & Landau,S. (2010). *Parents Perceptions of Their Role in School Leadership.* Poster presented at the annual conference of the National Association of School Psychologists, San Francisco, California.

Invited Talks and Lectures

Abney, D.H. (2016). *Complexity matching, coordination patterns, and development.* CogNetwork Colloquium Series, University of California, Berkeley, Berkeley, CA.

Abney, D.H. (2015). *Complexity matching, coordination patterns, and development.* Department of Psychological and Brain Sciences, Indiana University, Bloomington, IN.

Abney, D.H. (2014). *Complexity matching in conversational contexts.* CogNetwork Colloquium Series, University of California, Berkeley, Berkeley, CA.

Funding and Awards

2016	UC Merced GradSLAM Finalist ($250) University of California, Merced
2015	Hatano Cognitive Development Research Fellowship ($1000) University of California, Merced
2015	Cognitive and Information Sciences Summer Fellowship ($2,250) University of California, Merced

2014	Birnbaum Award Society for Computers in Psychology
2014	Cognitive and Information Sciences Travel Fellowship ($1,050) University of California, Merced
2014	Cognitive and Information Sciences Summer Fellowship ($3,800) University of California, Merced
2013	Cognitive and Information Sciences Travel Fellowship ($2,000) University of California, Merced
2013	Cognitive and Information Sciences Summer Fellowship ($2,500) University of California, Merced
2012	Charter Department Graduate Student Excellence Award Department of Psychology, Illinois State University
2011	Outstanding Psychology Graduate Assistant Award Department of Psychology, Illinois State University
2010	Rodney L. Custer Graduate Student Research Award Department of Psychology, Illinois State University
2010	The Donald Maley Spirit of Excellence Outstanding Graduate Student Citation International Technology and Engineering Education Association
2010	Graduate Student of the Year Department of Technology, Illinois State University

---

Professional Affiliations

2016–*present*	International Congress of Infant Studies
2015–*present*	International Environmental Communication Association
2014–*present*	Society for Computers in Psychology
2012–*present*	Psychonomic Society (Student Member)
2012–*present*	Cognitive Science Society (Graduate Student Member)
2011–*present*	Association for Psychological Science (Graduate Student Affiliate)
2011–2013	Midwest Psychological Association
2009–2010	International Technology and Engineering Educators Association

---

Teaching Experience

Instructor of Record

Fall 2009	Technology and the Quality of Life (TEC 275) Illinois State University, (Normal, IL)

Teaching Assistant

Spring 2016	Complex Adaptive Systems (COGS 180/204) Instructor of Record: Dr. Michael Spivey University of California, Merced (Merced, CA)
Fall 2015	Applied Ethics (PHIL 003) Instructor of Record: Dr. David Jennings University of California, Merced (Merced, CA)
Spring 2015	Introduction to Language and Linguistics (COG 005) Instructor of Record: Dr. Teenie Matlock University of California, Merced (Merced, CA)
Fall 2014	Speech Processing (COG 151) Instructor of Record: Dr. Anne Warlaumont University of California, Merced (Merced, CA)
Fall 2013	Mind, Brain, & Cognition (COG 101) Instructor of Record: Dr. Chris Kello University of California, Merced (Merced, CA)
Fall 2012	Mind, Brain, & Cognition (COG 101) Instructor of Record: Dr. Chris Kello University of California, Merced (Merced, CA)

Professional Service

2015	An Evening on Free Will (Co-organizer) University of California, Merced (Merced, CA)
2015	Second Annual Symposium on Child and Family. Professional growth event for community child educators University of California, Merced (Merced, CA)
2013–2015	Undergraduate Cognitive Science Student Association (CSSA) co-founder/representative University of California, Merced (Merced, CA)

2013	Student research talk organizer, Department of Cognitive and Information Sciences. University of California, Merced (Merced, CA)
2010	Department Chair search committee (student group), Department of Psychology. Illinois State University (Normal, IL)

Reviewer for the following journals

Attention, Perception, & Psychophysics, American Journal of Psychology, Developmental Science, Frontiers in Psychology, Journal of Experimental Psychology: Human Performance and Perception, New Ideas in Psychology, Perception, PLOS One, Behavioral Research Methods

Reviewer for the following proceedings

International Conference on Perception and Action (2015)
Annual Meeting of the Cognitive Science Society (2014, 2015, 2016)

Technical Skills

MATLAB, R, Python, Adobe ActionScript, Adobe Illustrator/Photoshop/Lightroom, SPSS, Superlab, PsychoPY, Excel, AutoCAD, 3D Studio MAX, LaTeX

Abstract

The study of human communication incorporates disciplines across the sciences and the humanities. One question that is important for better understanding and explaining human communication is how information is transmitted from one person to another person during an interaction. To communicate, humans produce and perceive complex behaviors such as vocalizations and body movements. Although researchers are beginning to better understand the production and perception of communicative behaviors, less work has focused on investigating the functions of these behaviors for information transmission during an interaction. Here, in collaboration with various co-authors, I present a hypothesis for human communication that has specific predictions for information transmission across individuals during an interaction.

The Complexity Matching hypothesis for human communication suggests that when the complex, hierarchical patterns of communicative behavior between individuals match, information transmission is enhanced. This hypothesis is motivated by work in statistical mechanics showing that when complex properties of two networks match, information transmission across the networks is optimal. In this dissertation, I present three projects that seek to test the Complexity Matching hypothesis for human communication.

First, I present initial observations of the production and convergence of hierarchical patterns of vocalizations during conversation. This study provides initial support for the Complexity Matching hypothesis and provides insights into the hierarchical properties of communicative behavior.

Next, I test the key prediction of the Complexity Matching hypothesis for human communication: enhanced information transmission. Pairs of adults were given a dyadic problem-solving task of building a tower structure out of a limited amount of materials. We observed that dyads built taller tower structures when their hierarchical patterns of vocalizations and body movements matched. These results provide initial support for the information transmission prediction of the Complexity Matching hypothesis.

Finally, I investigate the development of hierarchical structure in human communication. This study follows daylong vocal recordings of infants and their caregivers across the first two years of life. We observed evidence

for hierarchical patterns of vocalizations at the earliest recordings session (second week of life) and a dynamic trajectory of complexity matching and other vocal coordination patterns across development.

This dissertation, *The Complexity Matching Hypothesis for Human Communication*, is submitted by Drew H. Abney in 2016 in partial fulfillment of the degree Doctor of Philosophy in Cognitive and Information Sciences at the University of California, Merced, under the guidance of dissertation committee chair Christopher T. Kello.

Chapter 1

Introduction

As humans, we produce and perceive behavioral patterns by way of complex interactions between our brains, our bodies and its modalities, and the environment. Despite the prevalence of these rich complex patterns in our world, less is known about their underlying functions.

The study of human communication is a prime example of production and perception of behavioral patterns within and across people. For example, during a conversation with a friend about how to find a specific trail that leads to a secret climbing area, I will produce vocalizations and body movements such as gestures to communicate my thoughts. Similarly, my friend will perceive the patterns of vocalizations and body movements I produced, and produce his own to communicate whether or not he understands my directions, if he has some follow-up questions, or if he is nervous about the upcoming arduous hike.

There are decades of research devoted to understanding the patterns of vocalizations that span multiple levels of linguistic representations. There has been, perhaps, even more work dedicated to understanding movement patterns. Despite the long history of studying the production and perception of vocalization and movement patterns, only in the past decade have researchers sought to better understand multimodal patterns of behavior during communicative interactions.

Human communication research is multifaceted and interdisciplinary. My interest in diving into the vast field of communication research is to better understand the functions of complex behavioral patterns during human conversation and interaction. The key question is: what are the behavioral and coordination patterns that lead to optimal information transmission between two people in an interaction? Even a preliminary answer to this difficult question would have a significant impact in many intellectual arenas, from communication theory to the humanities, and from physics to developmental psychology.

There are many strategies for attempting to tackle the question of human information transmission in communicative settings. I could focus my attention on a specific level of linguistic representation, like phonetics, and investigate the phonetic patterns that lead to subjective and objective outcomes during human interaction. Such work has been done and has provided a rich account of patterns of convergence during vocal interaction (Pardo, 2006, 2013) and undoubtedly inspired others to think similarly about convergence and alignment of vocal patterns and linguistic representations during human communication. The strategy I will take in this dissertation is a more holistic perspective. Instead of focusing on a single level of linguistic representation or timescale of movement patterns, I will focus on the hierarchical, nested patterns of human behavior.

Perhaps the most accessible example of hierarchical patterns of human behavior is the production of human speech and language. 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 within the system scales across levels of measurement (Bak, 2013; Kello, Beltz, Holden, & Van Orden, 2007; Kelso, 1997; Van Orden, Holden, & Turvey, 2003; West & Deering, 1994). 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).

It is reasonable to assume that hierarchical structure of speech and language has function. After all, it is generally agreed upon that speech and language display hierarchical structure presumably reflecting how multiple linguistic representations are structured (Pickering & Garrod, 2004). When discussing hierarchical structure in communicative behavior, I am referring to the acoustic structure emerging from vocalization dynamics and the visual structure emerging from perceived body movements. These hierarchical patterns are less understood and correspond more to hierarchical patterns of low-level physical/acoustic energy.

In this dissertation, I am interested in understanding the production and convergence of hierarchical patterns of behavior during human communication. Uncovering new patterns and properties of human behavior is essential for progressing our understanding of human interaction. But merely uncovering new patterns and properties is not sufficient. Is a pattern or property of human behavior incidental? Or, is a pattern or property functional? Finding function is crucial for the progress of linking newly discovered patterns and properties of human behavior to the understanding of development, cognition, physiology, communication, and beyond. The key question posed within this dissertation is whether or not there is function to hierarchical patterns of behavior during human communication.

1.1 The production and convergence of hierarchical patterns of communicative behavior

What behavioral and coordination patterns lead to greater information transmission across two people in an interaction? Past studies established that, during dialogue, interlocutors match properties of phonetic productions (Pardo, 2006, 2013), speech pauses (Cappella & Planalp, 1981), syntactic structures (Bock, 1986), and lexical expressions of confidence (Fusaroli et al., 2012), etc. A natural progression is to consider if these various levels of linguistic representation produced by two individuals during an interaction become correlated with each other. Pickering and Garrod's (2004) interactive alignment model provides a framework for alignment within and across linguistic levels.

Indeed, Pickering and Garrod's interactive alignment model is a good starting point for understanding the relationships between levels of linguistic representation during dialogue, more specifically, and for building intuitions about a hierarchical description of communication more generally. However, similar to the deep and necessary question of convergence of a single level of linguistic representation, what is the function of hierarchical convergence? The null hypothesis for the function of hierarchical convergence is that the non-random convergence is simply an incidental property of the interaction.

One alternative hypothesis is that the function of the hierarchical convergence between two or more interacting systems is to facilitate information transfer. Classes of complex systems can be formalized statistically, relative to the dynamics of their interacting components. West et al. (2008) analyzed the coupling dynamics of complex systems in terms of the temporal clustering properties of their activity. The temporal clustering of events follow a power law, which is a precursor to hierarchical nested patterns. Analyses have shown that information transmission between coupled complex systems are maximal when the properties of their power laws are similar (Aquino, Bologna, Grigolini, & West, 2010; Aquino, Bologna, West, & Grigolini, 2011; Turalska, West, & Grigolini, 2011). This observation has been termed Complexity Matching and has since received interest from cognitive scientists and psychologists (Abney, Paxton, Dale, & Kello, 2014; Coey, Washburn, & Richardson, 2014; Coey, Washburn, Hassebrock,

& Richardson, 2016; Fine, Likens, Amazeen, & Amazeen, 2015; Marmelat & Delignières, 2012; Torre, Varlet, & Marmelat, 2013).

One approach for estimating the hierarchical structure of human behavior has been to quantify the temporal patterns of event series during human interaction (Abney, Kello, & Warlaumont, 2015). When events cluster across multiple temporal scales – a term called multiscale clustering – we can estimate the degree of nested hierarchical structure. For power laws in the multiscale clustering of point processes, convergence of exponents corresponds with convergence in the amount of temporal clustering across timescales. West et al.'s Complexity Matching (2008) provided theoretical motivation for expecting convergence in the temporal multiscale clustering of communicative behavior: Under these conditions, information exchange should be maximized between interlocutors as complex systems (Stephen & Dixon, 2011; Stephen, Stepp, Dixon, & Turvey, 2008).

1.2 The Complexity Matching hypothesis for human communication

The Complexity Matching hypothesis for human communication suggests that when the hierarchical structure of communicative behavior (e.g., speech or body movements) converge between two people, information exchange is enhanced. There is a growing field of research studying coordination patterns during human interaction (Dale, Fusaroli, Duran, & Richardson, 2014; Shockley, Santana, & Fowler, 2003; M. J. Richardson, Marsh, Isenhower, Goodman, & Schmidt, 2007; Schmidt & Richardson, 2008). Most of the research on coordination patterns during human interaction has focused on the temporal or distributional convergence between two people. For example, Paxton and Dale (2013) observed that the degree of temporal convergence of body movements during conversation depends on whether interlocutors are arguing about topics of disagreement or whether they are discussing topics like their favorite movies, songs, books. But what is the function of these convergence patterns? Louwerse et al., (2012) observed that multimodal temporal convergence between two people performing a challenging joint task strengthened (1) the longer the dyads conversed with each other and (2) when the task was more difficult. These studies provide descriptions of convergence patterns and their consequences during specific types of conversational and interactional contexts. However, more work is needed to go beyond description and provide theory-driven explanations of successful and unsuccessful communication.

The question of *why* temporal convergence occurs during human interaction has motivated a number of hypotheses that relate to the question of *function*. Pickering and Garrod (2004) suggested that perhaps the function of multilevel alignment of linguistic representation is the convergence of shared representations which can be considered communication. The shared representations build 'common ground' (Clark & Marshall, 1981), which is understood as a precondition for successful communication (Pickering & Garrod, 2004).

Another hypothesis is that human interaction, and specifically, interpersonal coordination, can be considered a coordinative structure (Bernstein, 1967). A coordinative structure is a self-organized set of interacting components that emerge under certain constraints. Researchers have applied the notion of coordinative structures to interlimb rhythmic coordination (Fuchs, Jirsa, Haken, & Kelso, 1996), interpersonal interlimb coordination (Schmidt, Carello, & Turvey, 1990), and interpersonal coordination across modalities and task constraints (M. J. Richardson et al., 2007; Shockley, Baker, Richardson, & Fowler, 2007; M. J. Richardson, Marsh, & Schmidt, 2005).

Pursuing the hypotheses of shared representations and coordinative structures during human interaction has greatly advanced our understanding of communication, coordination, and human interaction. However, these hypotheses do not have clear predictions for information transmission during human interaction. Complexity matching, on the other hand, has a clear prediction for information transmission of interacting complex systems.

Recent empirical efforts applying the concept of complexity matching to human interaction have 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 re-analysis of speech signals similar to Abney 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.

1.3 The present work

My research in the past few years has focused on quantifying patterns of human behavior and interaction that can be diagnostic to (1) varying conversational contexts, (2) successful and unsuccessful information transfer, and (3) the development of vocal communication. This dissertation focuses on better understanding the hierarchical patterns of human behavior during human communication. I have proposed the Complexity Matching hypothesis for communication, which suggests that information transmission is enhanced when the hierarchical structure of communicative patterns converge between two people during an interaction.

The first section presents the initial observation of the matching of multiscale clustering of human speech during conversation. Most research studying human interaction and coordination has focused on local temporal patterns of matching behavior like postural sway (Shockley et al., 2007), body movement (Richardson et al., 2007), and language (Fusaroli et al., 2012). This first study provides evidence for complexity matching in human conversation and shows that hierarchical patterns of vocalizations during conversation and the degree of matching of these patterns between interlocutors vary as a function of conversation type. This study is our introduction to the notion of hierarchical patterns of communicative behavior in human communication: the multiscale clustering of vocalizations.

The next section looks to test the hypothesis that information transmission is enhanced during human interaction when the hierarchical patterns of communicative behavior converge. In this study, I focus on the multiscale clustering of vocalizations and body movements during a dyadic problem-solving task. The Complexity Matching hypothesis suggests we should expect that higher rates of complexity matching between the vocalizations and movements of dyad members should lead to better performance on the problem-solving task. I provide evidence that complexity matching of both vocalizations and body movements are associated with increased performance on a dyadic problem-solving task. This study provides preliminary support for the Complexity Matching hypothesis.

The final section investigates the development of the hierarchical structure of human communication. Up until this point, I have focused on the production and convergence of multiscale clustering of adult human behavior across interaction contexts. However, there are many different coordination patterns that span various time scales and levels of description. In order to better understand the complex patterns of human communication, this study focuses on investigating the relationship between different vocalization coordination patterns across development. This study follows the daylong vocal recordings of fifteen infants and their caregivers across the first two years of life. Not only do I observe dynamic patterns of complexity matching between infants and caregivers, I also observe that infants at the youngest age of recording show complex, hierarchical structure in their vocalizations. These observations provide additional insights into the origins and development of communicative structure.

Chapter 2

Complexity matching in dyadic conversation

2.1 Preface

In this chapter, I will present a published study providing evidence for complexity matching in human adult conversation. The study re-analyzed vocalizations from a dataset of different dyadic conversations (Paxton & Dale, 2013a). In the study, participants conversed about different topics that constrained the conversations into affiliative and argumentative conversation types. The results showed that multiscale clustering of vocalizations differed as a function of conversation type. Additional results provided evidence for complexity matching and that the degree of complexity matching differed as a function of conversation types. These results provide initial evidence for complexity matching in human interaction and show the context-specificity of the production and convergence of multiscale properties of speech.

2.2 Introduction

Conversation is a complex coordination of human behavior (Shockley, Richardson, & Dale, 2009). Interlocutors need to attend to each other flexibly and continuously over the course of conversation so that they know what to say and when to say it in order to, if successful, satisfy their conversational goals.

One prominent model of dyadic conversation is Pickering and Garrod's (2004) interactive alignment model. The model emphasizes the importance of aligning different linguistic representations between interlocutors and predicts that two people in a conversation match representations at different linguistic levels. There are numerous schemes for dividing linguistic processing into levels, but Pickering and Garrod (2004) discuss six: phonetic, phonological, lexical, syntactic, semantic, and situational. In support of this model, a range of studies has shown that interlocutors match speech behaviors at various scales of linguistic structure. Interlocutors have been shown to match productions of phonemes (Pardo, 2006), speech pauses (Cappell and Planalp, 1981), syntactic structures (Bock, 1986), and descriptive utterances (Garrod & Anderson, 1987). In these cases, there are direct correspondences between particular instances of behaviors, such as mimicking individual utterances, syntactic phrasings, accented words, and so on. We shall use the term behavioral matching to refer to these phenomena alternately known as alignment, entrainment, convergence, and synchronization (Louwerse, Dale, Bard, & Jeuniaux, 2012).

A growing body of literature supports the existence of behavioral matching, but the specifics and interpretation are matters of debate. Some argue that behavioral matching and related processes are integral to dyadic interactions (Pickering & Garrod, 2004), while others emphasize the role of behavioral matching in facilitating mutual comprehension (Brennan & Clark, 1996). Others argue that principles and processes of perception and action give rise to behavioral matching (M. J. Richardson et al., 2007; Sebanz, Bekkering, & Knoblich, 2006). Still others contend that human communication is a general framework for situated action in which interlocutors maximize detection and sensitivity to others (Suchman, 2007).

These ongoing debates have been useful and informative because they suggest that behavioral matching plays some role in establishing common ground and, more generally, facilitating communication. However, opportunities for behavioral matching in natural conversation are limited because interlocutors do not simply mirror each other's behaviors. Each person makes unique, individual contributions to dyadic interactions, but effective communication necessitates that interlocutors share common ground and coordinate behaviors (Healey, Purver, & Howes, 2014; Mills, 2014). Thus many aspects of conversational behavior may be expressed by more indirect, subtle forms of coordination. Even turn-taking is more complex than synchronization or syncopation. Turns often do not alternate cleanly and evenly (Stivers et al., 2009), and interlocutors often speak and gesture simultaneously during periods of so-called "back channeling" (McClave, 2000).

The irregular, complex nature of dyadic interaction raises the question of whether behavioral matching may be generalized to more indirect forms of matching. That is, the drive to establish common ground and facilitate communication may be addressed through other means that can be viewed as extensions of behavioral matching. One natural extension is distributional matching—the idea that behaviors may match at the level of statistical, ensemble characterizations, rather than the level of particular behavioral acts. For instance, mean speech rates may converge during conversations (Webb, 1969), or two interlocutors may converge in their proportions of slang expressions, without directly matching each other slang for slang. The concept of distributional matching is consistent with Pickering and Garrod's (2004) interactive alignment model. Perhaps the best example comes from the well-known phenomenon of syntactic priming (Bock, 1986; Pickering & Branigan, 1998, 1999), in which hearing or seeing the usage of a given syntactic form (e.g., active vs. passive) increases the likelihood that speakers will use it themselves. Syntactic priming can arise from behavioral matching or distributional matching. In the latter case, the probability distributions over syntactic forms may converge between interlocutors (Jaeger & Snider, 2008).

The hypothesis of distributional matching takes on a new dimension when the distributions being matched follow power law functions (Clauset, Shalizi, & Newman, 2009). A power law function expresses one variable as a nonlinear function of another variable raised to a power, , where, . The heterogeneities and irregularities of language behaviors are reflected in many different power laws – frequencies of word usage and rank, (Zipf, 1949), frequencies of n-grams in text corpora (Kello & Beltz, 2009), frequencies of syntactic links to words (Ferrer i Cancho, Solé, & Köhler, 2004), correlations and burstiness across vowels/consonants, letters, words, and topics (Altmann, Cristadoro, & Degli Esposti, 2012), and spectral density of fluctuations in audio power of music and human speech (Voss & Clarke, 1978). These power laws reflect the heterogeneity of language in terms of variability across a wide range of measurement scales. They correspond to the irregularity of language in terms of rough stochastic patterns, unlike the highly regular fractals (i.e., power laws) of snowflakes and Mandelbrot sets. Below, we provide a descriptive example of a power law distribution in language and outline a method for its estimation.

In the present study, we find evidence for a new power law distribution in conversational speech signals. The power law is hypothesized to reflect hierarchical clustering and levels of linguistic information in conversational speech (Grosjean, Grosjean, & Lane, 1979), akin to levels proposed for the interactive alignment model. The speech data come from dyadic conversations designed to be either affiliative or argumentative (Paxton & Dale, 2013a), and the speech signals are analyzed in terms of their temporal dynamics, as captured by acoustic onset events and subsequent periods of acoustic energy.

In the present study, the power law in event clustering is measured by the Allan Factor (AF) function, which computes coefficients of variation across multiple timescales. Measured AF functions are found to converge in dyadic conversations, particularly for affiliative conversations and not argumentative conversations. We call this convergence complexity matching as a special case of distributional matching when distributions are power laws. The term comes from studies in statistical mechanics (West, Geneston, & Grigolini, 2008) showing maximal information exchange between coupled complex systems that individually produce similar power laws.

We explore whether conversational speech signals exhibit the conditions predicted from statistical mechanics on the approach that complexity matching can provide a unique angle into naturalistic conversation. We compare behavioral and complexity matching to test whether they make distinct contributions towards explaining dyadic interaction, and whether complexity matching yields useful evidence beyond behavioral matching.

2.3 Power law clustering in conversational speech

A simple way to approximately describe a power law distribution is to say that variability occurs across a wide range of measurement scales, including timescales. For the latter, imagine that a time series of measurements is windowed and the average measured value is computed for each window of size S. Variability across scales means that measures of variance scale up with window size S, e.g. small variations for millisecond windows, larger variations over seconds, even larger variations over hours, and so on. Variability across scales is unexpected for most types of simple systems. For instance, if one measures the temperature fluctuations in a refrigerator, variations would actually decrease with larger time windows, because larger windows would yield averages that converge on or near the temperature setting.

Variability that spans measurement scales is indicative of power laws, and such power laws will emerge from more complex systems, namely ones which display hierarchically nested structures and processes (Simon, 1977). In particular, sentences are collections of syntactic phrases, phrases of words, words of syllables, syllables of phonemes, and so on. Such nested levels of linguistic representation are integrated in the interactive alignment model, as illustrated in Figure 1. We expect the hierarchical nesting of language to be physically manifested as power laws in speech signals.

Hierarchical nesting in speech signals can be illustrated as follows. At the coarsest timescales, when two people converse, each interlocutor produces turns – long, clustered periods of acoustic speech energy interspersed with mostly no acoustic energy while the other person is talking. At finer timescales, there are breaks in the signal due to thinking time, phrase boundaries, rhetorical effects, and the like. At still finer timescales, breaks occur sometimes at word boundaries, and sometimes at phonemes with little or no sonority, such as plosive consonants (e.g., p,t,k,b,d,g), quiet

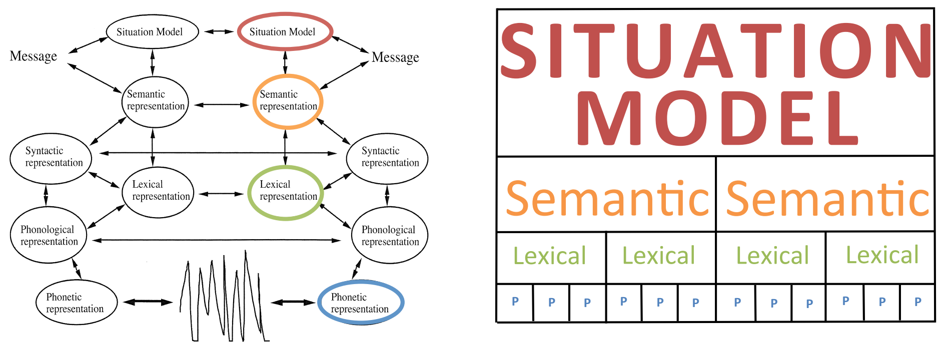

Figure 2.1: (Left) Pickering and Garrod's (2004, reprinted with permission) schematic representation of the stages of comprehension and production processes according to the interactive alignment model. (Right) An illustration of the nesting of different scales of linguistic representations, using four levels of the six from the interactive alignment model: phonetic, lexical, semantic, and situation model.

fricatives (e.g., f,h,th), and even voiced fricatives and nasal stops in some cases (e.g., v,m,n,ng). All of these breaks are defined as falling below some threshold of acoustic energy, i.e., we do not assume total silence or even a total lack of perceptible sound during breaks.

The three illustrative scales just listed are visualized in the speech waveform displayed in Figure 2. It is important to note that one could posit additional or different scales as well. Whatever the case, their physical manifestations are likely to overlap and blend such that one simply observes clusters of acoustic energy across a continuous range of scales in the raw speech signal. In fact, a continuous range of scales is expected to emerge when interactions propagate across levels of representation (Mitzenmacher, 2004; Holden, Van Orden, & Turvey, 2009), as posited in the interactive alignment model. Phonetic processes interact with lexical processes, which interact with syntactic processes and feedback to phonetic processes, and so on.

2.4 Complexity matching in speech signal clustering

Our discussion so far leads us to expect power law clustering in speech signals due to the hierarchical nesting of language representations and processes. Thus we need a method for measuring and quantifying clustering in speech signals across different timescales. Clustering is expected specifically in the timing of periods of acoustic energy interspersed with breaks as defined by some threshold. Such temporal clustering can be measured in the onset times when acoustic energy crosses from below to above threshold. Acoustic onset times are not only appropriate for measuring temporal clustering, but they also are highly salient and important events in speech perception (Cummins & Port, 1998; Cutting & Rosner, 1974; Liberman, Harris, Hoffman, & Griffith, 1957). Clustering in acoustic onset times is visible in Figure 2.

The interactive alignment model holds that interlocutors "align" representations across levels of linguistic processing. The particular nature of alignment is an ongoing area of research, and as mentioned earlier, behavioral matching is one manifestation of alignment that is well-documented in the literature (Louwerse et al., 2012). But also as mentioned earlier, behavioral matching is limited because direct correspondences alone cannot explain the rich behavioral diversity in natural conversations (Healey, 2008; Howes, Healey, & Purver, 2010).

Temporal clustering of acoustic onsets across scales, as a physical expression of linguistic processing across levels of representation, affords the possibility for a kind of distributional matching distinct from behavioral matching. The overall amount of temporal clustering can be quantified as a function of timescale, as we explain more formally below. Conversational speech signals may converge in terms of the distribution of temporal clustering across timescales. Such convergence would constitute a complex coupling in the dynamics of linguistic processing. This coupling would be complex partly because it would go beyond synchronization and other simple phase relations between time series, and partly because it would constitute the coupling of two power law distributions that reflect nested, interactive scales of processing.

Power law distributions are defining of complex systems in general (Sales-Pardo, Guimera, Moreira, & Amaral, 2007; Simon, 1977). Specifically, a complex system is one in which microscopic events may cascade up to alter macroscopic patterns of activity, which in turn may constrain and shape its microscopic events (Stanley, 1987). By this definition, both humans and human languages are demonstrably complex systems (Beckner et al., 2009; Mitchell, 2009; Kugler, 1987; Spivey, 2007; Swenson & Turvey, 1991; Thelen & Smith, 1994). Molecular and cellular events cascade up to affect behavior via myriad genetic and physiological processes, and behavior helps shape those processes via evolution and learning, for example. Likewise, microscopic changes in phonetic features may alter entire words, sentences, and conversations as macroscopic patterns, and the latter provide higher-level constraints on how phonemes are phonetically realized.

Classes of complex systems can be formalized statistically, relative to the dynamics of their interacting components. West and colleagues (West et al., 2008) recently analyzed the coupling of complex systems in terms

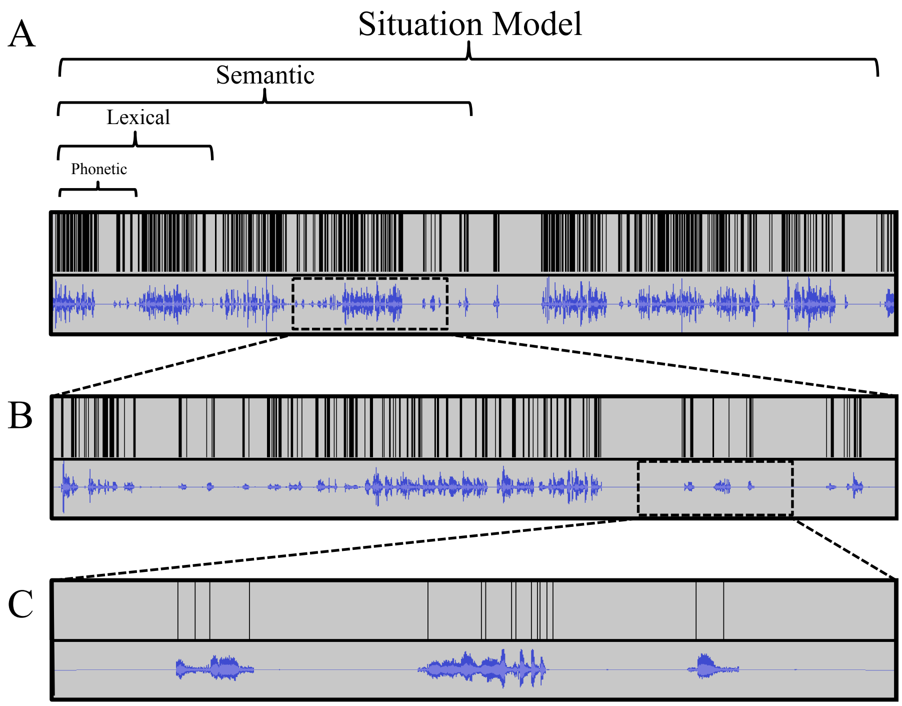

Figure 2.2: (Left) An example conversational speech signal, shown at three different temporal scales. (A) The longest scale roughly corresponds with conversational turns. The phonetic, lexical, semantic, and situation model labels approximate the time scales of these units on the speech signal. (B) The middle scale roughly corresponds with e.g. thinking pauses and phrase boundaries. (C) The shortest scale roughly corresponds with word, syllable, and phoneme boundaries. Vertical lines show acoustic onsets relative to a threshold chosen by visual inspection.

of their event dynamics, which amounts to temporal clustering of point processes analogous to acoustic onsets. Interestingly, analyses have shown information exchange between coupled systems to be maximal when the exponents of their power laws are similar (Aquino et al., 2010, 2011; Turalska et al., 2011) . For power laws in the temporal clustering of point processes, convergence of exponents corresponds with convergence in the amounts of temporal clustering across timescales. Thus West et al. (2008) provide independent theory and rationale for expecting convergence in the temporal clustering of conversational speech signals—under these conditions, information exchange should be maximized between interlocutors as complex systems (Stephen et al., 2008; Stephen & Dixon, 2011).

The formal analysis conducted by West et al. (2008) relies on statistical physics and mechanics and its elaboration is outside the scope of the current article. However, we can draw an intuitive analogy with simple oscillators designed to illustrate coupling beyond synchronization. Imagine two metronomes whose kinematics are coupled through a physical medium such as a sliding platform (Figure 3a). Provided that their frequencies are sufficiently similar, and coupling is sufficiently strong, the beats of the metronomes will tend to synchronize over time (Strogatz & Mirollo, 1991; Kelso, 1981). The phase-coupled oscillations that result from these interacting forces can be seen as idealized forms of behavioral matching, and a number of dyadic interaction studies have drawn this parallel (Schmidt & Richardson, 2008).

Now imagine two sets of metronomes at each end of the platform (Figure 3b) whose resonant frequencies span a wide range of timescales, and do not correspond one-to-one across the two ends of the platform. Coupling may still yield a system for which synchronization is an inherently low energy state, but synchronization and other simple phase relations may no longer be sufficiently strong attractors to create stable dynamical states of the system. This is more likely to be true especially when coupling is relatively weak. In such cases, the system instead is prone to exhibiting intermittent, irregular transitions from one metastable state to the next (Kelso, 1997). Such complex dynamics are readily observable in systems as simple as coupled oscillators, and coupled oscillators provide only a simple model of human interlocutors. Thus the metronomes serve to illustrate how complex couplings are not exotic or rare, but rather, quite expected for interactions between such richly heterogeneous systems like humans.

Expectations of phase couplings and more complex couplings lead us to predict behavioral matching and complexity matching in human interactions. To our knowledge, this prediction has not been tested previously for conversational interactions, but we can find support for a similar hypothesis in human perceptual-motor interactions (Coey et al., 2014; Marmelat & Delignières, 2012). Marmelat and Delignières (2012) recently conducted

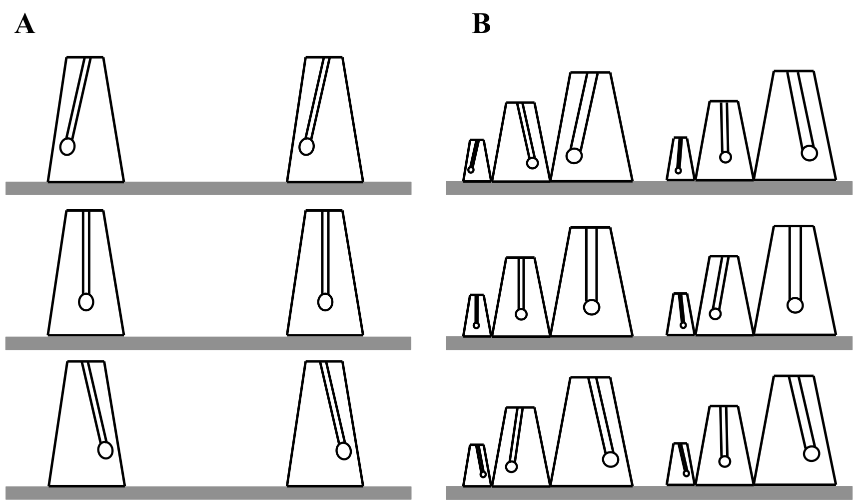

Figure 2.3: (Left) Examples of synchronization and behavior matching with toy metronome systems. (A) Illustration of two metronomes interacting along a sliding platform, as a simple model of synchronization and a form of behavioral matching. (B) Illustration of interactions between multiple metronomes with differing frequencies, to aid the intuition of complexity matching.

an experiment in which each participant in a dyad swung a hand-held pendulum, with instructions to swing in synchrony. Synchronization is a form of behavioral matching, but deviations from synchrony were analyzed for power law fluctuations in the form of $1/f$ alpha noise. Results showed that alpha estimates for each member of a dyad were correlated to the extent that coupling was facilitated by visual and physical contact. These alpha correlations served as a direct measure of complexity matching, and they could not be explained in terms of behavioral matching because dyadic time series of deviations from synchrony were uncorrelated at all lags – i.e., there were no cross-correlations.

2.5 Current Study

The new contributions of the current study are tests of (1) power law clustering in the temporal patterning of acoustic onsets in conversational speech and (2) complexity matching in the temporal clustering of speech across timescales. Power law clustering is expected to manifest due to the hierarchical nature of language processes. Complexity matching is expected to extend and complement behavioral matching, as part of a broader basis for interactive alignment that enhances communication through increased information exchange.

Our study was designed to investigate complexity matching through a number of different conditions and analyses. First, we analyzed data from a recent study by Paxton and Dale (2013) in which participants who previously did not know each other were asked to have two conversations (order counterbalanced). One was a casual, affiliative interaction about popular media. The other was on provocative issues based on participants' closely held beliefs and designed to evoke more argumentative conversations. Beforehand, participants were given questionnaires to gauge their opinions on these provocative issues, and specific issues were chosen if participants had strong but differing opinions about them. Partners were instructed to converse for ten minutes in each condition, which provided ample time for long stretches of speech to be analyzed. The original aim of the study was to investigate alignment in asymmetric contexts, that is, interactions between interlocutors who have conflicting, differing, or opposing goals and opinions.

These experimental data serve the current goals quite well, because the time series that can be extracted from audio data are long enough to afford measurements of temporal clustering across a wide range of scales. In addition, we can test for a relationship between complexity matching and a high-level discourse constraint: conversation type. Testing for such a relationship is important for providing converging evidence that temporal clustering of acoustic onsets is reflective of levels of linguistic processing rather than just matching of low-level acoustic properties of speech. The experiment also allowed us to compare matches between two speech signals from an originally paired dyad, with mismatches between signals from two different dyads. The latter provides a baseline for measuring complexity matching above chance and is a common baseline among dyadic interaction researchers (Bernieri, Reznick, & Rosenthal, 1988).

Another important feature of the experiment by Paxton and Dale (2013) is that it allows us to compare our measure of complexity matching with a more traditional measure of behavioral matching, where the latter can be quantified through cross-correlations in speech signals. As elaborated below, greater behavioral matching in our case corresponds to the negative peak of the cross-correlation function, which reflects the complementary turn-taking relationship between the temporal patterns of acoustic speech energy produced by each member of a dyadic conversation. We directly test whether complexity matching can be reduced and attributed to behavioral matching as measured by negative peaks in cross-correlations, or whether the two reflect distinct aspects of coordination in dyadic conversation.

2.6 Methods

2.6.1 Participants

A total of 28 undergraduate students (mean age=20.14 years; females=22) from the University of California, Merced participated in return for extra course credit. Individual participants signed up for time slots anonymously, and participants were not informed of their partner's identity beforehand. Dyads included 8 female-female and 6 mixed-sex pairings. By chance, male-male pairings did not occur. All participants reported conversational fluency in English and normal or corrected hearing and vision. Participants also reported their native language as either English (n=10), Spanish (n=10), or other (n=6; two participants did not disclose their native language).

2.6.2 Procedure

Before conversing with one another, each participant completed a brief series of questionnaires, including an opinion survey on political, social, and personal topics (e.g., abortion, death penalty, gay/lesbian marriage, legalization of marijuana). For each topic, participants were asked to write a brief synopsis of their opinion and mark how strongly they held their opinion from 1 (feel very weakly) to 4 (feel very strongly) on a Likert-style scale. Experimenters determined the topic of argument by comparing the two participants' survey answers to identify the topic on which participants held strong but opposing views. This topic was chosen as the dyad's argumentative prompt, given along with an instruction to convince one another of their opinion. Two additional prompts were also selected by those criteria but were given only if the participants were unable to continue the conversation on the topic at hand. Of the 14 dyads analyzed here, 10 required additional prompts (secondary=9; tertiary=1). In addition to the argumentative conversation, each dyad also had an affiliative conversation. The affiliative prompt instructed each dyad to identify and discuss popular media that both participants enjoyed. Affiliative prompts were designed to emphasize the common ground between partners, whereas the argumentative prompts were designed to emphasize their differences of opinion. Following the questionnaires, participants were brought together in a private room and seated facing each other. To provide an opportunity for partners to become acquainted with each other, they were left alone for about three minutes to introduce themselves outside the context of the experiment, without yet knowing the nature of the experimental task. To make introductions as natural as possible, participants were told that experimenter had to step out of the room to complete last-minute paperwork before beginning the experiment. After the introduction period, the experimenter entered the room and delivered the first conversation prompt. The order of prompts was counterbalanced across dyads, and participants were not informed of upcoming prompts. During each 10-minute conversation, the experimenter monitored recording equipment from a seat on the periphery of participants' range of vision. After each conversation, participants were separated and asked to complete post-conversation questionnaires. At the end of the experiment, participants were thanked and debriefed.

2.6.3 Apparatus, data collection, and data preparation

Conversations were recorded on a Canon Vixia HF M31 HD Camcorder, mounted on a Sunpak PlatinumPlus 600PG tripod. Audio for each participant was recorded separately at 44 kHz sample rate, using an Azden CAM-3 mixer and Audio-Technica ATR 3350 lapel microphones affixed to the upper portion of each participant's shirt. Two audio files were recorded per conversation (one for each interlocutor), which yielded four files per dyad and 56 files altogether across the 14 dyads.

After truncating audio files to contain only the conversations, Audacity was used to remove non-speech signals, as well as any partner cross-talk so that each file contained only one participant's speech signal. The Audacity "sound finder" was then used to locate acoustic onset and offset events in each file. The signal/no-signal threshold of acoustic intensity was set at -30db for all dyads, which was judged to be the lowest threshold that resulted in less than approximately 5 percent spurious onset events. This threshold yielded an average of 764 paired onset and offset events per partner, per conversation. For every audio file, the resulting event time series was highly irregular and clustered, based on visual inspection. Each event series was unique, as expected given that each partner made unique contributions to their conversations. However, we are interested in statistical quantities that abstract away from particular event times and characterize their temporal properties.

2.7 Inter-event intervals

The interactive alignment model, along with its hierarchically nested levels of linguistic processing, leads us to predict complexity matching in the temporal clustering of acoustic onset events. However, West et al. (2008) showed that complex systems in general are expected to exhibit complexity matching when their *inter-event intervals* (IEIs) are power law distributed with an exponent near two, $P(IEI) \sim 1/IEI^{\gamma}$, where $\gamma \sim 2$. West and colleagues' analysis suggests that we test IEIs for the predicted power law.

A histogram of IEIs was computed for the time series from each participant in each conversation, where the position of the smallest bin was set relative to the shortest IEI value in each given time series. The nine subsequent bins were logarithmically spaced to capture IEIs of all lengths for each time series. Logarithmic spacing accounted for the anticipated power law in IEI distributions—that is, greatest resolution in the histogram is needed for at the small end of the scale because the vast majority of IEIs are relatively short, and resolution can become coarser as IEIs become larger and less frequent.

Figure 4 shows the resulting histograms for each participant, plotted together in a single graph. Plotting individual histograms together provides a picture of the overall trend of the distributions, as well as the individual variability around that trend. The figure shows a clear trend of a negatively sloped line in logarithmic coordinates that flattens out for the shortest IEI values on the left. The slope of the trend is about -2 for both conversation types, as can be seen by comparing with the dashed line which has a slope of exactly -2. Thus the data closely resemble the theoretically derived precondition for complexity matching , i.e., the power law, $P(IEI) \sim 1/IEI^{\gamma}$, where $\gamma \sim 2$.

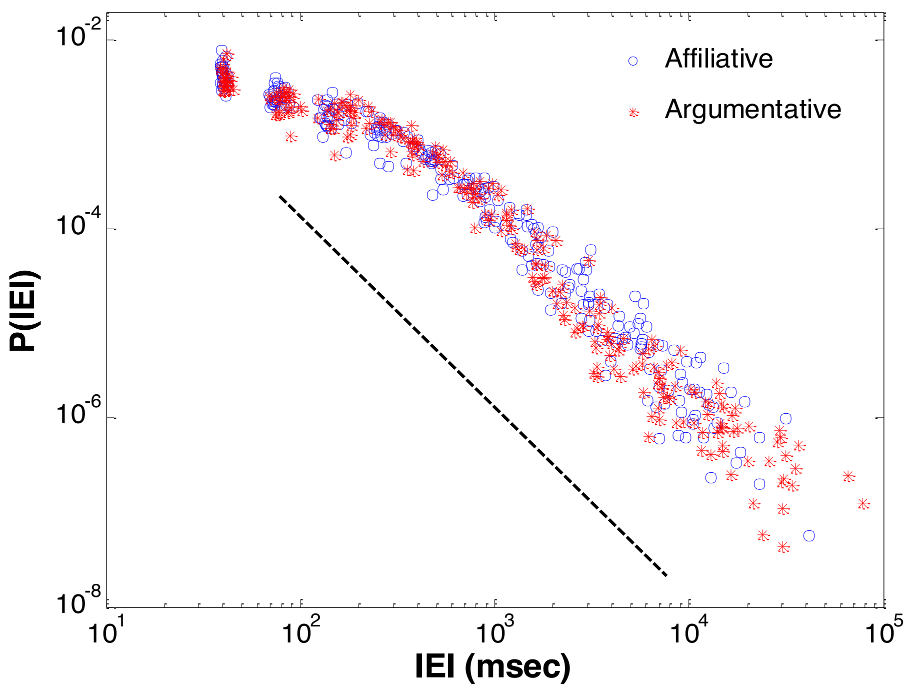

Figure 2.4: (Left) IEI probability density functions for individual interlocutors in individual conversations, plotted in logarithmic coordinates using logarithmic binning. Dashed line shows idealized slope of -2 (per West et al., 2008).

2.8 Temporal clustering in acoustic onsets

To quantify temporal clustering in acoustic onsets and test for a power law across timescales, we adopted Allan Factor (AF) analysis that has been used to measure temporal clustering in neural spike trains (Teich, Heneghan, Lowen, Ozaki, and Kaplan, 1997). Spikes and acoustic onsets are both examples of *point processes*, i.e., time series of events treated as occurring at instantaneous points in time. A Poisson process is one whose events occur unpredictably through time, i.e., for which knowledge of any and all event times up to a given point in time t provides no information about when future events may occur. AF is a statistical method that distinguishes between Poisson processes and those whose events occur non-randomly. In our case, we are interested in non-Poisson processes whose events cluster at different timescales more than would be expected by a Poisson process.

AF analysis is partly illustrated in Figure 2. Time series are tiled with adjacent windows of given size T; in the figure, each bracket represents one window of a given size. Events are simply counted within each window, and a measure of variance—AF variance, $\mathrm{A}(T)$—is derived from the differences in counts between adjacent windows. $\mathrm{A}(T)$ is calculated for a range of window sizes (i.e., timescales T), and Poisson processes are those for which $\mathrm{A}(T)$ approximately 1 for all T. Clustering at a given scale results in $\mathrm{A}(T)$ > 1, and more specifically, clustering across scales means that $\mathrm{A}(T) \sim T^{\alpha}$, where $\alpha \sim 0$. Finally, complexity matching is measured as the difference between two $\mathrm{A}(T)$ functions, where more matching corresponds to smaller differences.

2.9 Formal description of AF analysis

A formal description of AF analysis is as follows. A given point process is segmented into M adjacent windows of size T (enough to span the entire series), and 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 $\mathrm{d}(T) = N_{\mathrm{j+1}}(T) - N_{\mathrm{j}}(T)$. $d(T)$ values are computed for each of a range of values for T. The AF variance $\mathrm{A}(T)$ for a given timescale T is the expected value of the squared differences, normalized by mean counts of events per window (i.e., a type of coefficient of variation)

$$A(T) = \frac{\left\langle d(T)^2 \right\rangle}{2 \langle N(T) \rangle} \tag{2.1}$$

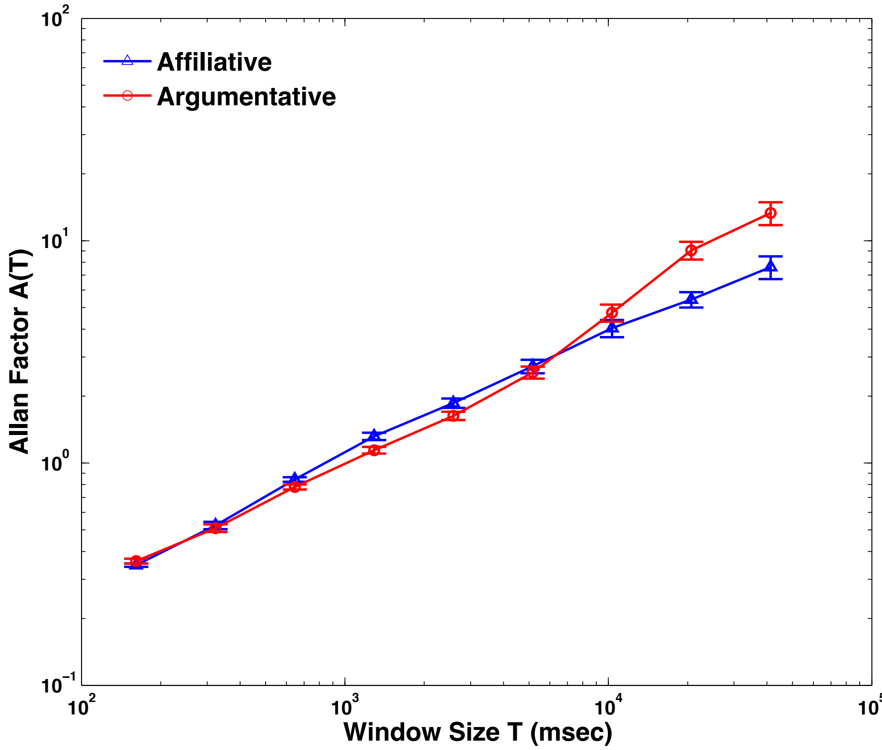

Figure 2.5: (Left) Mean AF functions for argumentative vs. affiliative conversation types, with standard error bars.

Poisson processes yield $A(T) \sim 1$ for all T, whereas power law clustering yields $A(T)$ approx. $(T/T_1)^\alpha$, where T_1 is the smallest time scale considered and alpha the exponent of the scaling relation. Point processes with $\alpha \sim 0$ are Poisson-distributed, whereas power law clustering means meaning $\alpha > 0$ over the measurable range of timescales (Thurner et al., 1997).

$A(T)$ was computed for each event time series from each interlocutor in each conversation. Each time series was 10 min long, and time windows varied as a power of 2, $T = 2$t where t ranged from 4 to 12. The resulting timescales ranged from 160 ms to 41 s. Smaller timescales were excluded because they are heavily affected by measurement error, and larger timescales could not be reliably estimated given the length of time series. $A(T)$ values were averaged across participants for each conversation, and averages are plotted as a function of T in Figure 5.

2.10 Results of AF analysis

A clear power law is evident in the roughly linear relationship in logarithmic coordinates for both conversation types. This power law is evidence of nested clustering of events over the measured timescales, as expected for nested language processes. The exponent of the AF power law was estimated for each individual time series by taking the slope of a regression line fit to each AF function in logarithmic coordinates. Mean exponent estimates for affiliative conversations (M = 0.53, SE = .02) were reliably less than those for argumentative conversation (M = 0.63, SE = .02), $t(27) = 4.57$, $p < .001$. This effect can be seen in Figure 5 as deriving from A(T) differences at the largest timescales. In general, this is evidence that the clustering of acoustic onsets reflects linguistic processing during conversations, rather than purely acoustic structure.

More specifically, results showed greater temporal clustering of onsets in argumentative conversations relative to affiliative ones, at longer timescales. Longer timescales mainly reflect turn-taking dynamics, which suggests that there were fewer, longer turns in argumentative conversations. To confirm this interpretation of the observed difference in AF functions, we compared the number and mean duration of IEIs greater than four seconds. Four seconds was approximately where the AF functions diverged, and was a cutoff that should capture mostly turn intervals, i.e., an utterance without a break in acoustic energy, followed by a pause before the partner begins the next turn. We did not expect this automated method to capture turns perfectly—some turns will be missed or cutoff, and some intervals will reflect utterances within turns—but it is safe to assume that the majority of these few very long intervals (less than 5 percent of all intervals on average) mostly correspond with turns. As expected, estimated turns for argumentative conversations were found to be fewer (M = 21.7 versus M = 26.8, $t(27) = 2.9$, $p < .01$) and longer (M = 12.6 versus M = 8.0, $t(27) = 5.2$, $p < .001$), compared with affiliative conversations.

2.11 Complexity matching

The previous two sections established two preconditions necessary to test for complexity matching, i.e., (1) power law distributions in IEIs that approach an exponent of two and (2) power law clustering of acoustic onsets, as expressed in the AF function, that ostensibly reflects the hierarchical nesting of linguistic processing during conversation. Now, to test for complexity matching, we need a measure of similarity between two AF functions and a baseline for the amount of complexity matching expected by chance. Our measure of AF similarity is the summed absolute difference between two AF functions a and b, with a negative log transformation:

$$D_{a,b} = -\sum log\left|A\left(T_a\right) - A\left(T_b\right)\right| \tag{2.2}$$

The log transformation takes into account the scaling law over T, and the negative simply makes greater values correspond with greater complexity matching, relative to a baseline control.

For baseline controls, we used surrogate comparisons between event series. Specifically, condition controls were created by comparing event series of two interlocutors from the same conversation type (either both affiliative or both argumentative) but who did not converse with each other. Fifty-two condition controls were created for each dyad in each condition, and the resulting Da,b values were averaged for each dyad. Mean Da,b functions are plotted in Figure 6 for original pairings and condition controls, separated by conversation type.

$D_{\mathrm{a,b}}$ values for original pairings in the affiliative conversation (M = 11.91, SE = 1.40) were greater than their condition controls (M = 9.80, SE = .44), $t(13)$ = -1.95, $p_{\text{one-tailed}} < .05$. However, there was no such effect in the argumentative conversation, $t(13)$ = -.06, $p_{\text{one-tailed}}$ = .48. These results provide evidence for complexity matching in the power law clustering of acoustic onsets in affiliative conversations, but not argumentative conversations. A qualitative inspection showed that AF differences generally occurred across timescales between affiliative originals and controls. Thus matching did not vary significantly over the range of timescales in which phonological, lexical, syntactic, and discourse processes unfold.

Finally, we note that the effect of conversation type was so strong that complexity matching for affiliative controls was a little more than that for argumentative original pairings, albeit not reliably so, $t(13)$ = 1.59, p = .134. The reason for this result needs further investigation, but one possibility is that argumentative conversations create a repelling dynamic that opposes complexity matching, thereby making speech signals no more similar than chance. This possibility is supported by analyses of behavioral matching reported next.

2.12 Behavioral matching

The previous section reported evidence for complexity matching, but it is important to test whether this evidence can be attributed to behavioral matching. Interlocutors' speech signals may exhibit "align-able" patterns in their periods of acoustic energy, possibly with some temporal lag between the signals. Phase-shifted alignment would constitute behavioral matching, and if the patterns are power law clustered, the same signal similarity that yields behavioral matching would also yield complexity matching. Here we

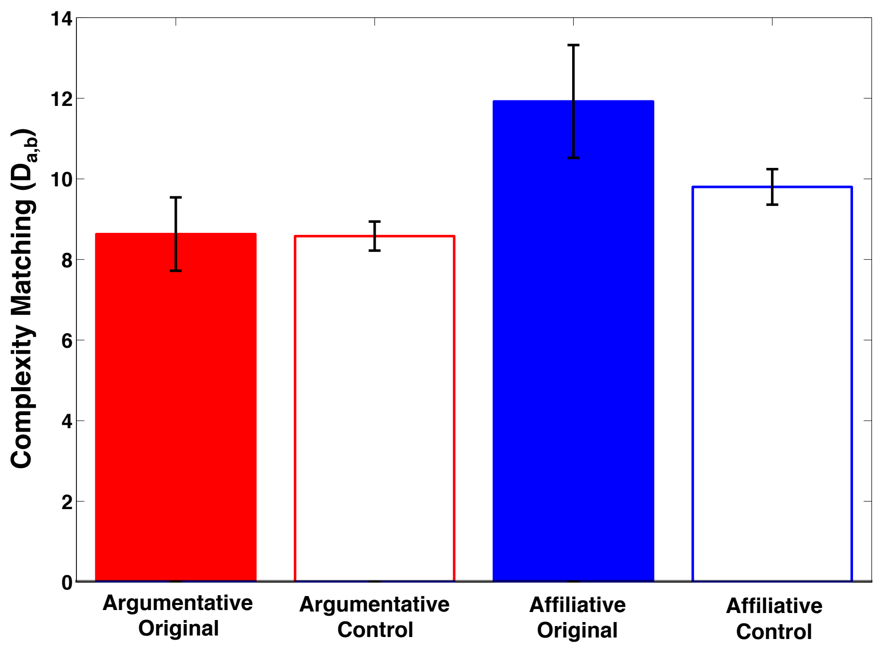

Figure 2.6: (Left) Mean summed AF difference functions plotted for the two conversation types, separately for original pairings versus randomized controls, with standard error bars.

test for behavioral matching in conversational speech signals and compare with complexity matching results to determine whether the complexity matching results may be attributed more simply to behavior matching. The most directly related measure of behavioral matching would be to use the same time series of acoustic onsets as used for complexity matching and simply cross-correlate them. However, point processes are theoretically instantaneous, which means their lack of duration complicates direct use of the cross-correlation function. Rather than assign each onset a temporal range, we used the duration of ongoing acoustic energy that followed each acoustic onset, i.e., the periods of acoustic energy from each onset to each subsequent offset. The resulting time series of acoustic energy periods were then cross-correlated to test for evidence of behavioral matching. Surrogate cross-correlation functions were also computed using the same method as that for AF functions.

Cross-correlations yielded no evidence of alignment at any lag: Peak positive correlation coefficients for affiliative pairings (M = .06, SE =.006) were not reliably different from their surrogate controls (M = .06, SE =.001), $t(13)$ = -1.14, p = .274, and the same was true for argumentative pairings (M = .09, SE =.01) compared with their surrogate controls (M = .07, SE = .002), $t(13)$ = -1.67, p = .119. These null results provide an initial suggestion that our complexity matching results cannot simply be attributed to behavioral matching.

Inspection of the cross-correlation functions reveal that, unlike peak positive correlations, peak negative correlations are far greater in magnitude for original pairings compared with surrogate pairings. This effect holds for both affiliative and argumentative conversations, $t(13) = 4.77$ and 5.92 (respectively), both $p < .001$. These negative peaks reflect complementarity in the time series of acoustic energy periods, which likely derives from turn-taking in conversational speech. Thus maximal misalignment is also a kind behavioral matching, albeit one where each speech act is matched with a lack thereof. While this may be considered as behavioral mismatching, it is demonstrative of a strict temporal coordination between partners that is very much in line with the spirit of behavioral matching research. This turn-taking measure of behavioral matching was stronger for argumentative conversations (M = -.32, SE = .16) compared with affiliative conversations (M = -.23, SE = .14), $t(13)$ = 4.16, $p < .001$. Thus the effect of conversation type on behavioral matching was different and opposite from its effect on complexity matching: Behavioral matching results point to stricter turn-taking in argumentative conversations, whereas complexity matching highlights stronger coupling across levels of linguistic processing in affiliative conversations. However, it is possible that our measure of complexity matching is somehow the converse of our measure of behavioral matching. If so, the two measures should be negatively correlated. Results did not bear out this hypothesis: A correlation of $D_{\mathrm{a,b}}$ values with peak minimum cross-correlations yields a coefficient of $r(28)$ = .23, which is not reliable, p = .243. This null result suggests that the complexity matching we observed cannot be straightforwardly attributed to behavioral matching.

2.13 General Discussion

Perhaps the most salient coordination we experience in conversations is behavioral matching. We take turns, echo speech acts of our partners, and strive for mutual understanding by sharing and in some sense matching our states of knowledge. The saliency of behavioral matching in firsthand experience has an analog in the scientific study of interpersonal coordination. Synchronization is a salient form of coordination dynamics, and one that is relatively easy to formalize and investigate mathematically (Schmidt, Morr, Fitzpatrick, & Richardson, 2012). Other phase relations—like anti-phase (Haken, Kelso, & Bunz, 1985; Keller & Repp, 2004) are also investigated in this area, but all such phase relations can be conceptualized as different types of behavioral matching.

If we introspect further into the nature of conversational interactions, we can find other more indirect forms of coordination in speech. The "tone" of a conversation, for instance, is not just carried by particular matches between turns, words, or other speech acts. Tone can be partly expressed as an approximate statistical convergence in, for instance, pitch, loudness, and pace of speech (Manson, Bryant, Gervais, & Kline, 2013; Neumann & Strack, 2000; Webb, 1969). Similarly, regional accents and dialects can be considered as a kind of convergence (Coupland, 1980) in the temporal dynamics of speech over multiple timescales and partly stem from common allophonic variations that are coordinated among populations of speakers over countless conversations.

In this study, we introduced complexity matching to the interpersonal interaction literature. Complexity matching was imported from West et al. (2008) to measure broad, statistical forms of coordination in conversational speech. By analyzing data from naturalistic conversations, we found that complexity matching provides a new window into interpersonal coordination beyond behavioral matching. We measured temporal dynamics in speech as expressed through clustering of acoustic onset events across timescales. We chose this measure in part because it is a purely temporal index of speech dynamics—each acoustic onset varies only in time and nothing else—and in part because it expresses temporal dynamics across timescales, from phonetic to lexical to turn-taking variations in speech timing.

Using AF analysis, we found evidence for multiscale dynamics in the power law clustering of acoustic onsets, as measured by the AF function, and we found greater clustering at longer timescales for argumentative conversations, as measured by greater AF exponent estimates. This effect of conversation type on AF exponents indicates that multiscale clustering reflects more than just low-level acoustic properties of speech. It also indicates that argumentative conversations are more structured at the larger timescales of turn-taking, and this interpretation is supported by cross-correlation analyses indicating stricter turn-taking in argumentative conversations.

While argumentative conversations show stricter turn-taking, only affiliative conversations demonstrated complexity matching, i.e., convergence in multiscale clustering. We interpret this difference as reflective of the more subtle forms of coordination in speech that we mentioned earlier. When people engage in affiliative interactions to converge on some mutual understandings and opinions, this convergence can be reflected in subtle aspects of their speech dynamics that operate similar to constructs like tone, pace, and style. AF analysis of acoustic onsets was able to capture such subtle aspects of convergence.

The present findings also are consistent with previous multimodal analyses of the conversations. As mentioned above, herein we found no evidence of complexity matching in argumentative conversations, yet there was more behavioral matching compared with affiliative conversations, as measured by peak negative cross-correlations. Consistent with this difference, analyses of movement dynamics also found no behavioral matching during argumentative conversations (Paxton & Dale, 2013a). In the future we plan to work on complexity matching analyses that may be applied to both movement and speech dynamics, in order to investigate whether multimodal coordination may further illuminate the coupling of interlocutors during affiliative conversations, and lack thereof during argumentative conversations.

2.13.1 Complexity matching and theories of conversation, coordination, and development

As discussed in the Introduction, our interpretation of complexity matching is consistent with the interactive alignment model (Pickering & Garrod, 2004). Language systems and processes are inherently multilevel, i.e., multiscale, and the interactive alignment model posits coupling across levels. The concept of complexity matching is exactly this—a kind of coupling across the scales of two interactive systems. The concept comes from work in statistical mechanics (West et al., 2008) that connects to the idea of interactive alignment. Multiscale systems with interactive levels of processing generally are expected to exhibit power laws as signatures of the complexity that is concomitant in such cross-scale interactions. West and colleagues report formal analyses to show that two multiscale, complex systems are most responsive to each other when their power laws converge, particularly near a specific exponent in the power law distribution of inter-event intervals.

We find it somewhat remarkable that data from dyadic conversations fit the theoretical predictions of a theory from statistical mechanics that was formulated for a very broad class of physical systems. In the current study, we found the estimated exponents of inter-event intervals during conversations indeed were near the predicted exponent value of two. Thus our study is an example of how work from statistical mechanics can inform and enhance specific theories (such as interactive alignment) in the psychological and cognitive sciences, and how interdisciplinary research can yield new disciplinary insights.

To illustrate this point further, formal analyses of complexity matching yield another theoretical prediction that has been pursued in other behavioral studies. As noted earlier, complexity matching is predicted to correspond with increased information exchange between two complex systems. The experiment analyzed herein did not include a direct measure of information exchange, but Fusaroli and colleagues (Fusaroli, Abney, Bahrami, Kello, & Tylén, 2013) re-examined data from a joint perceptual decision-making task (Bahrami et al., 2010) in which dyads collaborated on visual discrimination judgments. Speech signals were analyzed using similar methods to those herein, and measures of complexity matching were found to correlate with increased performance derived from joint decision-making. These results suggest that greater increased complexity matching can correspond with enhanced joint decision making. This might also relate to work suggesting that mutual comprehension is facilitated by behavioral matching (Brennan & Clark, 1996; Brennan & Hanna, 2009).

Thus far, we have discussed complexity matching primarily in the context of the interactive alignment model, but it is also related to theories of interpersonal synergy (Ramenzoni, Riley, Shockley, & Baker, 2012; Riley, Richardson, Shockley, & Ramenzoni, 2011) that have grown from synergies as theorized in motor systems (Bernstein, 1967; Turvey, 1990, 2007). Synergy is the emergence of coordination via reduction in the degrees of freedom in a system of many interacting components. This concept can be extended from physical systems into the interpersonal interaction domain. By extension, Fusaroli, Raczaszek-Leonardi, and Tylén (2014) proposed that interpersonal synergies should 1) be highly sensitive to conversational context; 2) adapt flexibly to changing needs of the task, and 3) self-assemble to minimize variance to manage the degrees of freedom within the interaction (Riley et al., 2011). These entailments of synergies may lead to specific, testable hypotheses about functional specificity and reciprocal compensation in interpersonal coordination.

Processes of perception and action have been argued to give rise to behavioral matching. For example, interpersonal postural coordination has shown to be constrained by social (Shockley et al., 2003), articulatory (Shockley et al., 2007), and visual (Giveans, Pelzer, Smith, Shockley, & Stoffregen, 2008) dynamics. Whether the approach to understand conversation and interpersonal coordination entails shared representations (Sebanz et al., 2006) or coordinative structures (M. J. Richardson et al., 2007), these proposals only investigate coordinative behaviors at a singular timescale. The current study supports the notion that conversation and interpersonal coordination can also be investigated using theory and tools emphasizing multiple scales of analysis, furthering the notion that language and conversation are hierarchically scaled phenomena. Therefore, complexity matching might entail processes of perception and action, however, it is very plausible that the same multiscale exploratory dynamics found to constrain perception (Stephen, Arzamarski, & Michaels, 2010; Stephen & Hajnal, 2011), might also constrain the perception and action of conversation and interpersonal coordination. Future work should consider the relationships between research studying the multiscale behaviors of agents interacting with the environment and research studying the multiscale behaviors of agents interacting with other agents.

Complexity matching and the application to conversation also relate to work in language and situated action (Suchman, 2007). Most generally, this work emphasizes the role of situations – and therefore, context – to constrain actions. This is in contrast to the general proposal that plans constrain actions. The current study showed that conversation type constrained the AF exponents of the clustering of speech signals, thereby constraining action in the form of speech production during conversation. The results herein support the notion that the situation, or context, constrains action.

Finally, complexity matching may be a fundamental building block for communication and social interaction across the lifespan. Evidence and theory have led developmental researchers to hypothesize that rhythmic coordination between infant and caregiver, akin to behavioral matching, is supportive of infant language learning (Feldman, 2007; Jaffe et al., 2001). But like speech in adult conversations, there is evidence that infant vocalizations also are organized into hierarchical clusters during typical development (Lynch, Oller, Steffens, & Buder, 1995; Oller, 2000). Recent results using AF analyses revealed complexity matching between infant prelinguistic vocalizations and caregiver speech (Abney, Warlaumont, Oller, Wallot, and Kello, in preparation). Taken together, these studies suggest that complexity matching may be foundational to the learning and development of interpersonal coordination and communication.

2.14 Conclusion

Interaction research has relied on measures of behavioral matching as a measure of interpersonal coordination for decades. Complexity matching is a new, complementary measure of coordination. Behavioral and complexity matching provided unique insights into the different interactions that occur during affiliative versus argumentative conversations—arguments had stricter turn-taking, whereas friendly conversations yielded distributional similarities that may reflect the establishment of common ground. Together, these analyses provide a richer view of interaction than either alone. These complementary analyses may be generalized and applied to yield similar insights in other areas of language and interaction research, wherever hierarchical nesting may yield power law scaling in the temporal dynamics of behavior.

Chapter 3

Multimodal complexity matching and information transmission in a dyadic problem-solving task

3.1 Preface

A key prediction of Complexity Matching is that information transmission between complex networks is optimal when complex behaviors converge. To test this hypothesis in human interaction, I developed a dyadic problem-solving task that serves as a method for quantifying indirect information transmission. Participants were instructed to build the tallest tower possible in fifteen minutes using only marshmallows and uncooked spaghetti. The problem-solving task provides sufficient constraints to increase communication between members of a dyad. The key hypothesis is that performance on the task – taller spaghetti-marshmallow tower – would relate to higher rates of convergence in the complexities of vocalization and movement dynamics. Results from this study will provide insights into the general hypothesis of complexity matching in addition to the patterns of communicative dynamics that lead to increased rates of communication.

3.2 Introduction

Problem solving with another person is an important and necessary skill. Understanding the dynamics that lead to successful and unsuccessful problem solving is important for theoretical developments of human interaction and communication as well as for applications for increasing efficiency and quality of life for people working in groups on a daily basis. Underlying successful problem solving is the expectation of sufficient information transmission across group members. In the current study, I investigate whether specific dynamic patterns of human behavior lead to successful problem solving, and by proxy, sufficient information transmission.

There is a long tradition of capturing the dynamics of interpersonal coordination across a diverse range of tasks. Among these coordination patterns, two distinct types have emerged: synchrony and complementarity. From interlimb rhythmic coordination (Schmidt et al., 1990) to postural sway (Shockley et al., 2003, 2007) and eye movements (D. C. Richardson & Dale, 2005), observations of a close temporal coordination between people has led many researchers to suggest that synchronous behavior might be a defining dynamic pattern of human interaction. However, in many contexts, 'doing the same thing at the same time' would lead to negative consequences. For example, in mature adult conversations, speech overlap is to be avoided (Stivers et al., 2009). Many interactional contexts require people to do similar actions that are not temporally aligned. Such *complementary* behavior has shown similar functional benefits as synchrony for many interactional contexts (Dale et al., 2014; Fusaroli & Tylén, 2015; M. J. Richardson et al., 2015).

Considering the two classes on coordination patterns known to be functional across interactional contexts, synchrony and complementarity, it is important to understand how multiple coordination patterns coalesce throughout interactions. It seems likely that group members use a variety of coordination patterns throughout an interaction. In the current study, I focus on understanding the relationship between multiple, nested patterns of behavior across interactional contexts and how these patterns relate to performance on a dyadic problem solving task.

Previous work has observed that the patterns of conversational speech can be quantified as hierarchical, nested communicative structures termed multiscale clustering (Abney, Paxton, et al., 2014). The multiscale clustering of interlocutors' conversational speech differed across argumentative and affiliative conversations. Importantly, the degree of matching of multiscale clustering between interlocutors differed as a function of conversational context. This matching was interpreted through the lens of a hypothesis in statistical mechanics termed Complexity Matching (West et al., 2008). Complexity Matching is the hypothesis that when the complexities of two complex systems match, information transmission between the systems is optimal. Similarly, Abney et al. observed preliminary evidence for Complexity Matching: the complexities of conversational speech between interlocutors matched in different ways depending on the conversational context.

The observation that the hierarchical, nested structure of speech matches or aligns between interlocutors can be connected to Pickering and Garrod's (2004) interactive alignment model. The model emphasizes the alignment of linguistic representations at multiple linguistic levels: phonetic, phonological, lexical, syntactic, semantic, and situational. The notion of hierarchical, nested patterns of behavior can be extended to other human behaviors like body movements. Instead of considering the synchronous or complementary behaviors of group members, we can look at how similar patterns of behavior between interlocutors match at various temporal scales.

Do dyad members produce similar movements patterns across multiple temporal scales? Similar to language, we can consider movement patterns to be hierarchical and nested (Gibson; 1979; Reed, 1996). Take for instance a conductor guiding an orchestra. A conductor's role is to guide the orchestra through a musical score using visible gestures from localized movements of the hands, to more gross body movements spanning the entire body. Similar to levels of nestedness of linguistic representations during a conversation, throughout an Act, a composer's movement patterns can occur at small timescales like cuing the string section with her baton to larger timescales like gesturing the tempo and meter across multiple measures.

In line with the example of the orchestra conductor, we might expect that people working together on a difficult task might produce similar movement patterns that span across multiple temporal scales. In this study, I focus on measuring the hierarchical, nested patterns of two modalities - vocalizations and body movements - for group members working together on a joint problem-solving task. Analogous to a functional hypothesis for synchrony or complementarity, I hypothesize that dyad members with similar hierarchically structured vocalizations and movement patterns will perform better on the problem-solving task. This hypothesis is motivated by the Complexity Matching hypothesis suggesting that when the complexities of two systems match, optimal information transmission occurs.

3.3 Method

3.3.1 Participants

Seventy-four undergraduate students from the University of California, Merced, participated in 36 dyads in return for extra course credits. Participants individually signed up using the anonymous online subject pool system and could not see their partners identity before arriving at the study location.

3.3.2 Materials and procedure

The materials and procedure were identical to what was reported in Abney, Paxton, Dale, and Kello (2015). Participants were asked to sit in one of two stationary chairs near a square table (76.2 cm by 76.2 cm W by 71.1 cm H). Seating arrangement was participant-initiated and experimenters were careful to not provide any explicit direction toward any of the two chairs. The two chairs and table were oriented such that the chairs were placed adjacent to each other, with the table rotated 45 degrees in line of sight of the camcorder (Canon Vixia HF M31 HD camcorder).

Once seated and oriented with a comfortable sitting posture, participants were outfitted with Shure Beta 54 super-cardioid microphone headsets. Participants were then instructed to construct the tallest tower structure possible within 15 mins using only the materials provided: one box ($\sim$10 oz) of large marshmallows and one box ($\sim$1 lb) of raw spaghetti. The participant seated on the right was only allowed to handle the marshmallows and the participant on the left was only allowed to handle the spaghetti. This constraint was imposed to increase the difficulty of the task and to stimulate interaction. Participants were not allowed to use partial or broken pieces of materials, and were instructed to remove any pieces of material that broke during the construction. Participants were instructed to remain seated during the task. Experimenters monitored the rules throughout the construction phase and reminded participants if they violated any rules.

Experimenters answered any questions participants had before beginning the task. An experimenter provided 5- and 1-min warnings. Once the time limit (15 min) expired, the experimenters recorded the height and weight of the tower, in addition to the amount of materials used in the final tower and the amount of broken materials discarded.

3.3.3 Vocalization and Movement analyses

Video files were truncated to contain only interactions occurring during the 15-min construction phase. For each truncated video file, an audio file (.wav) was extracted for the left and right channel, respectively, for subsequent audio cleaning and coding. A research assistant listened to each .wav file and omitted cross-talk and noise from non-communicative breaths. Each file was then analyzed using the Audacity "sound finder" to locate acoustic onset/offset intervals. The threshold of acoustic intensity was set at -30dB for all audio files (Abney, Paxton, et al., 2014). Acoustic intensity was not consistent across all audio files. Therefore, where appropriate, acoustic intensity was amplified by 6dB for an entire .wav file. Vocalization onsets were used in subsequent analyses.

The truncated video files were analyzed using a frame-differencing method

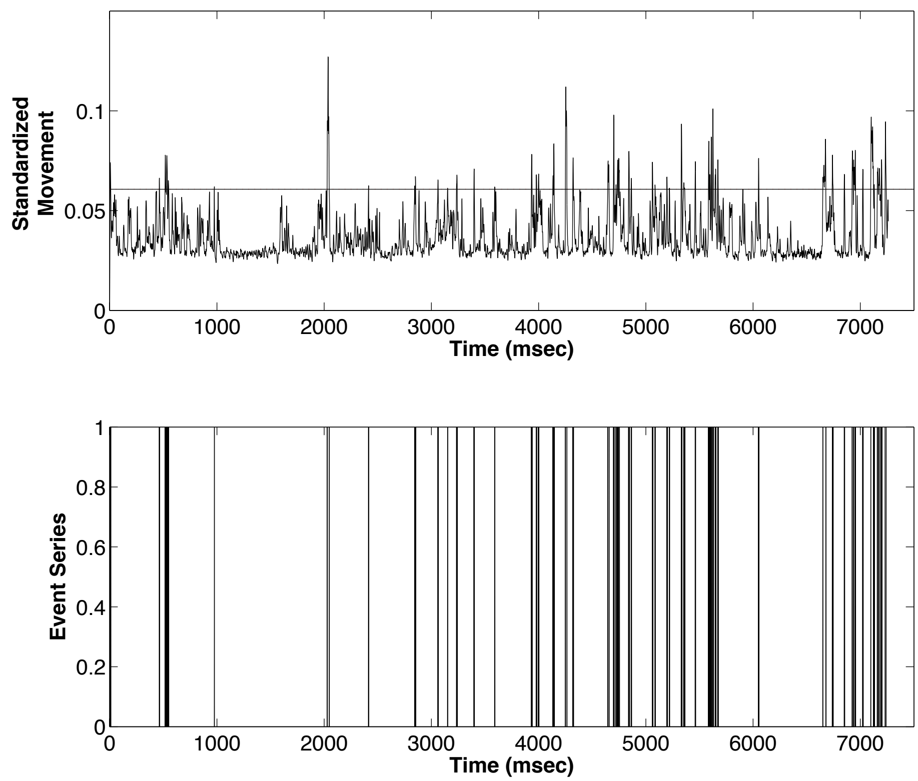

Figure 3.1: (Top) Example standardized movement difference series. Horizontal line represents event threshold. (Bottom) Example event series.

(FDM) to obtain time series of standardized movement indices computed from pixel-to-pixel changes from frame to frame (Paxton & Dale, 2013b). The FDM provided an objective measure of overall body movement. Higher FDM values indicated higher amounts of overall movement for each participant.

Binary spike trains of vocalization events were calculated by using the onsets estimated by the Audacity "sound finder". Vocalization onset events were coded as "1" and other states were coded as "0". Binary spike trains of movement events were calculated by creating a threshold of movement indices. A movement event was defined as a sample in the FDM movement series that exceeded 2 standard deviations above the mean movement index. Movement events defined by this threshold were coded as "1" and other states were coded as "0" (See Figure 3.1). Binary spike trains for vocalizations were downsampled to 8Hz to match the samping rate of the movement events. A multimodal spike train was also computed, combining vocalization and movement events. For each interlocutor, vocalization and movement spike trains were summed.

Allan Factor (AF) analysis (Allan, 1966) was used to estimate the multiscale clustering of vocalization and movement dynamics for each interlocutors' vocalizations and movement time series. The AF analysis estimated the variance of vocalization or movement events (e.g., onsets of vocalization or movement) at particular timescales and computed correlation estimates, α, across those multiple time scales. The AF analysis is a point process analysis that inputs a binary spike train of events (1s) and nonevents (0s) (see Figure 3.2).

$$A\left(T\right) = \frac{\left\langle d\left(T\right)^{2}\right\rangle}{2\left\langle N\left(T\right)\right\rangle} \tag{3.1}$$

To compute a measure of complexity matching of a dyad, I used an AF similarity metric introduced in Abney, Paxton, Dale, and Kello (2014). The AF similarity metric was the summed absolute difference between two AF functions, with a negative log transformation:

$$D_{a,b} = -\sum log\left|A\left(T_{a}\right) - A\left(T_{b}\right)\right|. \tag{3.2}$$

For each dyad, a complexity matching metric was computed for vocalizations, movements, and vocalization/movements.

The log transformation takes into account the scaling law over T, which is used to estimate the AF function. The negative sign makes larger values of $D_{a,b}$ correspond with greater complexity matching.

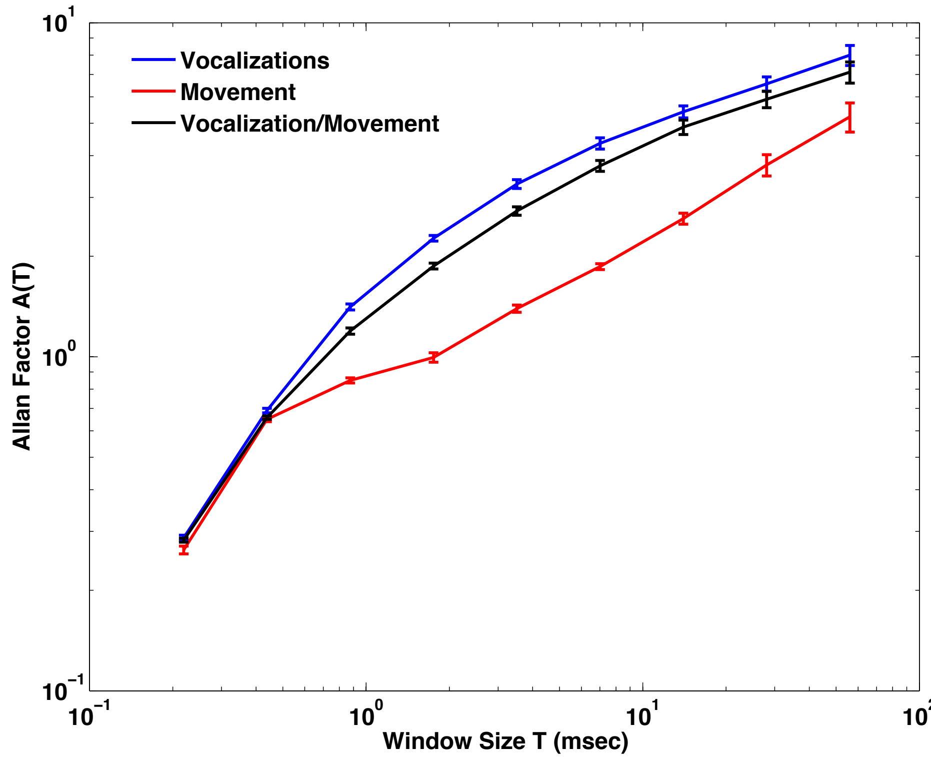

Figure 3.2: Mean Allan Factor functions for the Vocalization, Movement, and Vocalization/Movement time series. Error bars represent standard error.

3.4 Results

Average height of the spaghetti-marshmallow towers was 46.6 cm (18.36 in, SDcm = 15.40). To determine if complexity matching occurred beyond random pairings of interlocutors, I computed the average complexity matching scores for each participant and a random pairing of a participant who was not their partner. Empirical complexity matching scores for vocalizations (M=9.78, SD = 5.61) and movements (M=15.16, SD = 4.15) were significantly higher than surrogate complexity matching scores ($M_{\text{vocalizations}}$=6.74, $SD_{\text{vocalizations}}$ = 3.25,$M_{\text{movements}}$=11.47, $SD_{\text{movements}}$ = 4.12), ts(74) > 2.76, ps < .006.

Vocalization complexity matching significantly predicted tower height, b = .46, t(35)=2.87, p = .007. Vocalization complexity matching also explained a significant proportion of variance in tower height, R^2 = .17, F(1,35)= 8.21, p=.007. Movement complexity matching significantly predicted tower height, b = .38, t(35)=2.49, p = .02. Movement complexity matching also explained a significant proportion of variance in tower height, R^2 = .13, F(1,35)= 6.18, p=.02. Vocalization/movement complexity matching significantly predicted tower height, b = .47, t(35)=2.39, p = .02. Vocalization/movement complexity matching also explained a significant proportion of variance in tower height, R^2 = .12, F(1,35)= 5.71, p=.02. (See Figure 3.3).

It is possible that the relationship between complexity matching and tower height is driven by overall AF estimates. To test this possibility, for each modality, AF estimates were averaged across dyads and were submitted to correlational analyses with tower height. Averaged AF estimates for vocalizations (r[35]=-.19,p=.24) or vocalizations/movements (r[35]=-.26,p=.12) did not reliably correlate with tower height. Averaged AF estimates for movement did positively correlate with tower height, r(35)=.40,p=.01. Next, I residualized out averaged AF estimates for each modality with tower height and submitted the residual variables to regression models with the corresponding complexity matching metrics. Controlling for averaged AF estimates, vocalization complexity matching significantly predicted tower height (b = .40, t[35]=2.45, p = .02), whereas, movement complexity matching (b = .06, t[35]=.38, p = .71) and vocalizations/movement complexity matching (b = .32, t[35]=1.60, p = .12) did not predict tower height.

3.5 Discussion and Conclusions

In this study, I found that complexity matching occurred across vocalizations and movement patterns during a dyadic problem-solving task. The observations of vocal and movement complexity matching add to a

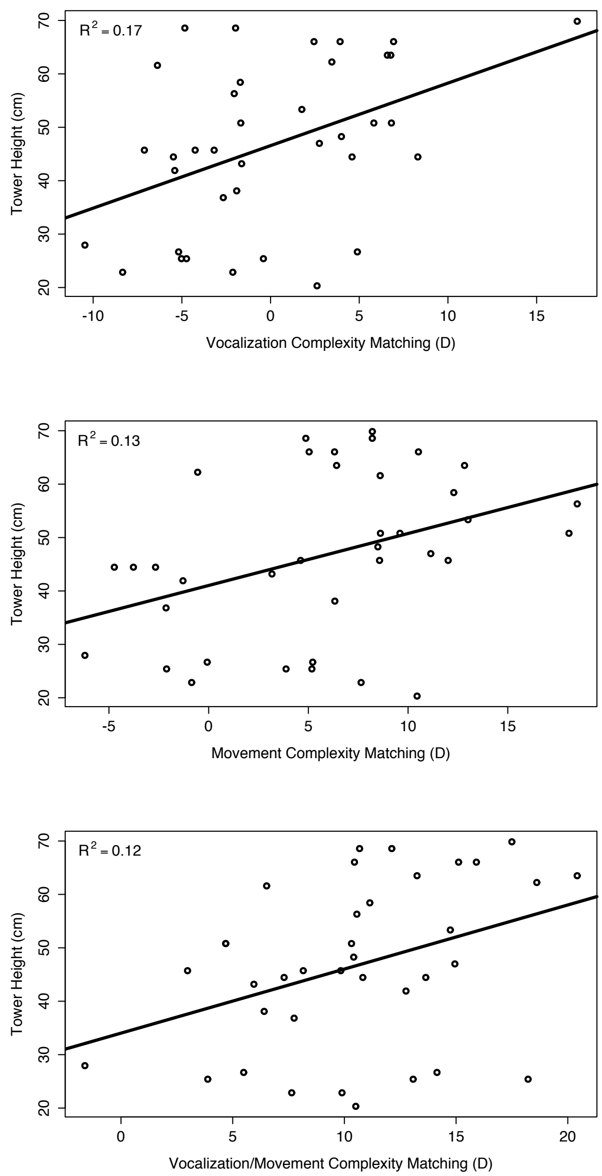

Figure 3.3: Complexity matching predicting tower height (cm). Individual dots correspond to each trial. Lines represent linear fit.

short, but growing list of conversational and interaction contexts that display descriptive differences in the degree of complexity matching (Abney, Paxton, et al., 2014; Coey et al., 2014, 2016; Fine et al., 2015; Marmelat & Delignières, 2012). The results presented here show evidence for the functional hypothesis of complexity matching: the degree of complexity matching predicted an indirect measure of information transmission.

This study also uncovered a novel coordination pattern: the matching of hierarchically structured movements across dyad members. In the vein of movement dynamics, there were three observations from this study that are new to the literature on coordination. First, I observed that overall body movement patterns are hierarchical and nested across multiple temporal scales. Second, dyad members' hierarchical patterns of movements matched more than surrogate dyads. Third, dyad members with higher rates of matching between their hierarchical movement patterns constructed higher towers. This final result is consistent with the Complexity Matching hypothesis: the matching between complexities of two systems related to increased information transmission.

Overall body movement patterns were hierarchically nested across timescales. Specifically, the onset patterns of movements were clustered across multiple temporal scales. To my knowledge, this is the first observation of this class of behavior in the literature on movement coordination. Shockley et al., (2003, 2007) observed that the postural sway between two people in a conversation matched in terms of the overall recurrence. The results in this study add nuance to local recurrence patterns of movements and suggest that there are clustered patterns of movements that occur across multiple nested temporal scales.

Higher degrees of matching between dyad members' hierarchical, nested pattern of vocalizations and overall body movements were associated with better performance on the problem-solving task. These results are consistent with the Complexity Matching hypothesis. When dyad members used similar vocalization and movement patterns spanning temporal scales, they constructed taller tower structures. It is likely this type of coordination pattern consists of both synchronous and complimentary patterns of behavior. Indeed, it is important to understand the contribution of multiple coordination patterns spanning levels of description and temporal and spatial scales. However, it is important to note that when controlling for overall hierarchical structure, only vocal complexity matching predicted tower height. One interpretation of this result is that vocal complexity matching matching was functional for the task, whereas movement complexity matching was an emergent coordination pattern, constrained by task properties, and therefore incidental to the task.

It is important to point out that a limitation of this study is the difficulty to define information and information transmission. The proxy for information transmission in the current study was performance on the task: higher tower structures. Of course, building tower structures does not directly equate to information transmission, but rather is an indirect measure of information transmission.

This study uncovered novel patterns of vocalization and movement coordination in a dyadic problem-solving task. A key observation was that vocalization and movement patterns could be quantified as hierarchical and nested across multiple temporal scales. Additionally, greater matching of hierarchical nested vocal and movement patterns between dyad members scale with better performance on a problem solving tasks. Future work should focus on understanding the contributions of multiple coordination patterns to successful and unsuccessful communication.

Chapter 4

Multiple coordination patterns in infant and adult vocalizations

4.1 Preface

In the last chapter, I provided preliminary evidence for a potential functional role for developing hierarchical structure in communicative behavior. One important question is how hierarchical properties of communicative behavior develop. Previous work has shown that hierarchical structure is present in mature adult vocalizations (Chapters 2 and 3), therefore it is possible that hierarchical structure is a vocalization property indicative of typical development. Furthermore, it is possible that the vocal interactions between infants and adults impact the development of important vocalization properties. In this chapter, I will present a study that investigated a large-scale corpus of naturalistic daylong recordings of infant and adult vocalizations across the first two years of life (Abney, Warlaumont, Oller, Wallot, & Kello, under review). The results show that the hierarchical structure of vocalizations produced by infants and adults match, even when controlling for other vocalization properties. Additional results suggest that the adults in the infants' auditory environments adapt the complexity of their vocalizations to be more similar to the infants' vocalizations over the first two years. Overall, the results provide evidence for complexity matching in early development and provide insights into vocal coordination and communicative development.

4.2 Introduction

The progression to speech-like vocalizations is a fundamental component of language learning (Oller, 2000) and is influenced by infant-adult vocal interaction (Bateson, 1975; K. Bloom, Russell, & Wassenberg, 1987; Goldstein, King, & West, 2003; Goldstein & Schwade, 2008; Jaffe et al., 2001; Kokkinaki & Kugiumutzakis, 2000; Nathani & Stark, 1996; Northrup & Iverson, 2015; Papoušek & Papoušek, 1989; Ramírez-Esparza, García-Sierra, & Kuhl, 2014; Warlaumont, Richards, Gilkerson, & Oller, 2014; Weisleder & Fernald, 2013). Likewise, the quality of these vocal interactions has been shown to predict other social and cognitive behaviors later in development. For example, in the seminal work by Jaffe et al. (2001), the authors found that the degree of vocal rhythmic coordination at four-months-of-age predicted levels of attachment and cognitive outcomes at twelve-months-of-age. There are additional studies exemplifying the notion that the degree of vocal interaction, either characterized in terms of temporal coordination (e.g., Feldman and Greenbaum, 1997) or in terms of other properties such as vocalization rate (e.g., Allely et al., 2013), predicts important developmental outcomes. Studies of vocalization properties and vocal coordination patterns are used to build theories of attachment (e.g., Bowlby, 1969) and social learning (e.g., Landry, Smith, & Swank, 2006) in addition to being markers of typical and atypical development (e.g., Oller et al., 2010; Patten et al., 2014; Warlaumont et al., 2014).

Finding new vocal coordination patterns and understanding the relationships between existing vocal coordination patterns might provide new insights into development. Furthermore, to advance our understanding on vocalization properties and vocal coordination, it is important to understand the similarities and differences between different measures of vocal coordination. In the current study, we investigate three different types of vocal coordination: coincidence-based, rate-based, and cluster-based. Coincidence-based and rate-based coordination have been previously used in a number of studies to study vocal interactions. Cluster-based coordination is a new measure recently introduced in the study of vocal interaction during adult conversation (Abney, Paxton, et al., 2014).

Coincidence-based coordination is based on the co-occurrence of vocalizations produced by two interlocutors within some minimal period of time. It includes both co-vocalizations (Harder, Lange, Hansen, Væver, & Køppe, 2015) and turn taking. Jaffe et al. (2001) observed that infant and adult vocalizations were contingent on each other up to a lag of 60s. They also observed that the strongest coordination patterns recurred every approx. 20s-30s (see also, Feldstein et al., 1993), which they suggested was the optimal interaction "rhythm". The degree of coincidence-based coordination was predictive of various measures of attachment and development. To quantify coincidence-based coordination in the present study, we used cross-recurrence quantification analysis to measure the degree to which infant and adult vocalizations occurred close together in time (Coco & Dale, 2013; Cox & van Dijk, 2013; Dale, Warlaumont, & Richardson, 2011; Fusaroli, Konvalinka, & Wallot, 2014; Marwan, Romano, Thiel, & Kurths, 2007; Warlaumont et al., 2014).

We based our measure of coincidence-based coordination on the timing of acoustic onsets of infant and adult vocalizations. Many previous studies have used similar measures of vocalization to study coordination. For example, van Egeren et al. (2001) found coordinated interaction within a temporal window of 3s between the onset of a vocalization produced by an infant and the onset of a vocalization response by the mother or vice versa (Harder et al., 2015). Akin to the measure of coincidence-based vocal coordination used in the present study, Warlaumont et al. (2014) observed that local coordination in timing of vocalizations across children and their caregivers differed as a function of vocalization type, and whether the infant was typically-developing (TD) or diagnosed with Autism Spectrum Disorder (ASD). Child speech-related vocalizations were more likely to receive an adult response relative to non-speech-related vocalizations. Children were also more likely to produce a speech-related vocalization if their previous speech-related vocalization received a response from their caregiver. Furthermore, relative to ASD children, TD children had more frequent vocal interaction with their caregivers and were more likely to lead vocalization interactions.

Rate-based vocal coordination is the degree of matching in the frequency or rate of a particular vocal behavior or property. One example of rate-based coordination is volubility matching. Volubility is the quantity or rate of vocalization per unit time, and volubility matching quantifies the similarity between infant and adult volubility across a given recording session. Much work has demonstrated volubility to be an important predictor of vocal development and communication (Franklin et al., 2014; Gilkerson & Richards, 2008; Goldstein & West, 1999; Goldstein, Schwade, & Bornstein, 2009; Hart & Risley, 1995; Hsu, Fogel, & Messinger, 2001; Oller et al., 2010; Oller, Eilers, Basinger, Steffens, & Urbano, 1995; Rescorla & Ratner, 1996; Warlaumont et al., 2014), but less work has quantified its coordination across infant and caregiver pairings. In one study, Hart and Risley (1999) found a positive relationship between infant and adult volubility. Other studies have examined effects of adult interactions more generally on infant volubility (K. Bloom et al., 1987; Franklin et al., 2014; Goldstein et al., 2009), and effects of adult volubility on child language learning (Ramírez-Esparza et al., 2014; Weisleder & Fernald, 2013), and cognitive and perceptual abilities (Greenwood, Thiemann-Bourque, Walker, Buzhardt, & Gilkerson, 2011; Jaffe et al., 2001).

Cluster-based vocal coordination measures the degree to which temporal events cluster similarly in infant and adult vocalizations. The acoustic energy in human vocalizations tends to be clustered in time (Abney, Kello, & Warlaumont, 2015; Abney, Paxton, et al., 2014; Luque, Luque, & Lacasa, 2015), in that there are frequent starts and stops due to many factors, including breathing, fluctuations in intensity, emotion, and so on. Clustering in speech vocalizations also emerges from variations in the sonority of phonetic features, prosody, and pauses due to thought and emphasis. Clustering in acoustic speech energy may also relate to the hierarchical clustering of linguistic units (Grosjean, Grosjean, and Lane, 1979): phonemes cluster into syllables, syllables cluster into words, words cluster into phrases, phrases cluster into sentences, sentences cluster into larger discourse-like structures, etc. (Pickering & Garrod, 2004).

Prelinguistic vocalizations, although not yet bounded by linguistic structure, show precursors to the hierarchical grouping of vocal events of mature speakers. For example, prelinguistic vocalizations produced by infants have been observed to follow a structure of hierarchical clustering at the grouping levels of syllables, utterances, and phrases (Lynch et al., 1995). Here we aim to extend this work by Lynch et al. by quantifying the degree to which infant vocalizations, and the adult vocalizations to which they are exposed, cluster across the day at timescales ranging from seconds to hours. In the present study, we investigate the developmental relationship between hierarchical clustering of temporal events in infant vocalization bouts versus adult vocalizations heard by infants, in addition to other vocal coordination patterns reflecting temporal-based and rate-based vocal coordination. It is generally accepted that the conversational exchange between interlocutors is a dynamic interplay with reciprocal effects (Snow, 1977) and that understanding how infants and adults modulate vocalization properties during conversational exchanges and across development is crucial for understanding typical and atypical communicative development.

In addition to investigating differences in the degree of coordination across the three levels of description (coincidence-based, rate-based, and cluster-based coordination), we can also investigate directions of convergence of these vocalization properties across infants and adults. For example, does volubility rate of caregiver vocalizations adapt to that of the infant? Or vice versa? Additionally, does clustering of caregiver vocalizations adapt to the hierarchical clustering of the infant?

4.2.1 Goals of the current study

The purpose of the current study was to examine the development of various types of vocal coordination across infancy and determine whether different patterns are interrelated or independent. We used the LENA™

(Language ENvironment Analysis) system (LENA Foundation, Boulder, CO) to collect naturalistic, daylong audio recordings from fifteen infants. The recordings are from an ongoing study in which infants are followed longitudinally during the first two years of life. The LENA system captures and automatically locates both infant and adult vocalizations. The present study seeks to answer three main questions revolving around the general theme of coordination patterns in vocal interaction: (1) Do coincidence-based, rate-based, and cluster-based coordination patterns vary depending on the vocalization type produced by the infant? (2) Do different coordination patterns provide unique information about the dynamics of vocal interaction? (3) How do the various coordination patterns relate to infant age?

4.3 Method

4.3.1 Participants

Participants were fifteen infants (7 females, 8 males) from an ongoing longitudinal study. Fourteen were exposed primarily to English and one was exposed primarily to German. The final analysis included 706 recording sessions; thus, the average number of recordings per participant was 47.06 recording sessions (SD = 11.53). The range of earliest recording session age was from 11 days-old to 162 days-old. The range of oldest recording session age was from 292 days-old to 885 days-old. Thus, the overall span in age range of infants was 11 days to 2 years; 5 months.

4.3.2 Data Collection

Recordings of infant and adult vocalizations were made using the LENA™ (Language ENvironment Analysis) system. The LENA system consists of an audio recorder that fits in the front pocket of custom-made clothing and a software system designed to automatically identify various speakers within the recordings (Xu, Yapanel, & Gray, 2009; Xu, Yapanel, Gray, & Baer, 2008). The automated system uses speech recognition technology, trained on human-annotated LENA recordings, to segment and identify onset times for specific vocalization types, taking into account the age of the infant (Xu et al., 2008, 2009). The procedure imposes a limit such that the minimal durations of an infant or adult vocalization segment are 600 ms and 1000 ms, respectively. Accuracy and reliability of the automated system has been tested against human transcribers for over 70 hours of American English data (Xu et al., 2009). Segment classification agreement between human transcribers and the automated system was 82% for adult vocalizations and 76% for infant vocalizations. For infant vocalizations, segment classification agreement between human transcribers and the automated system was 75% for speech-related vocalizations and 84% for non-speech-related vocalizations (Xu, Yapanel, and Gray, 2009). Timestamps of classified vocalization segments are reported in the LENA ITS (Interpreted Time Segments) file (Xu, et al., 2008). Infant and adult vocalization onset times were extracted from this ITS file (Warlaumont et al., 2014).

There are a few noteworthy limitations of the present study due to using the LENA system. Segments labeled as having overlap between an infant or adult and any other sound source, a very common label occurring in LENA automated analysis at all ages, were excluded because the system does not indicate the types of sound sources present in those segments. Although the excluded overlapping segments often include infant and/or adult segments, there is no way of knowing based on the automated labels when this is the case. There are also a number of factors that can potentially reduce the accuracy of classification. For example, environmental noise (Soderstrom & Wittebolle, 2013), infant age, speaker variation, and clothing type (VanDam, 2014) have been observed to influence accuracy (Xu et al., 2009). Our choice to use this system despite these limitations compared to human transcriptions is driven by the fact that the analysis of hierarchical structure of infant vocalization requires long time series in order to incorporate large temporal windows of analysis. The study described here would be impractical to conduct without automatic labeling of event onsets. Many of the same limitations also apply to a number of studies that have also used the LENA system to study language development (Ambrose, VanDam, & Moeller, 2014; Caskey & Vohr, 2013; Greenwood et al., 2011; K. Johnson, Caskey, Rand, Tucker, & Vohr, 2014; Oller et al., 2010; VanDam et al., 2015; Warlaumont et al., 2010; Warren et al., 2010); future studies and technological advances will be necessary to overcome these limitations.

The recorder captured each infant's voice as well as other sounds in the environment including adult vocalizations. In the present study, we utilized the automated speaker labeling provided by the software. Only timings of the onsets of each infant's vocalizations and of vocalizations produced by adults in the infant's proximal auditory environment were considered. We treated all recorded adult vocalizations, regardless of which particular individual produced them, as part of the same auditory stream, so when we refer to infant-adult interactions we are referring to the infant and all adults in the infant's auditory environment. Thus, our analyses do not distinguish between dyadic or triadic interactions where multiple adults are speaking. For the infant, vocalizations included speech-related sounds (e.g., babbling, singing, and gooing), reflexive sounds (e.g., cries and laughs), and vegetative sounds (e.g., burps and grunts). The vocalization onset times were obtained through a program that searched for onset times of CHN (i.e., Child) segments and AN (i.e., Adult) segments within the LENA ITS file. The program is available at https://github.com/HomeBankCode/lena-its-tools/releases/tag/v1.0 (Warlaumont, 2015).

Caregivers were instructed to begin recording when their infant awoke in the morning and to stop recording when their infant was put to bed at night. Audio recordings could be paused by the parents for privacy reasons throughout the recording sessions. If the caregiver paused and resumed recording in the same day, we treated each segment as a unique session.

1322 recordings sessions were collected across all infant-adult interactions. Recording sessions were omitted if the duration was less than 6hrs (505 session; 37.9 percent of original sample excluded), if the analysis of hierarchical structure could not converge due to low number of onsets (less than 200 onsets; 105 sessions; 7.9 percent), or if the estimate of hierarchical structure or volubility was 3.5 SDs above or below the mean (16 sessions; 1.2 percent). This left 706 sessions ($\sim$ 8492 recording hours) to be used in all the analyses reported below. Average session length was 12.03 hours (*SD*=2.72 hours). Sessions omitted due to the 6hrs criterion typically reflected the caregiver stopping the recorder at some point in the day and resuming recording at a later point.

For each session, four time-series of onset times were created, one for adult vocalizations, and three for the infant: speech-related (speech, non-word babble, and singing), non-speech-related (laughing, crying, burping, coughing, etc.), and the combination of speech-related and non-speech-related. These onset times served as the temporal events used to measure coincidence-based and cluster-based coordination.

4.3.3 Analyses

Coincidence-based coordination

To quantify the coincidence of infant and adult onset events, we used Cross-Recurrence Quantification Analysis (CRQA) to obtain a diagonal cross-recurrence profile (DCRP) (Coco & Dale, 2013; Dale et al., 2011; Warlaumont et al., 2014). A DCRP uses a 10s sliding window to assess overall quantity of coincidence-based coordination at a range of lags. Formal mathematical descriptions of CRQA and DCRP have been documented elsewhere (Coco & Dale, 2013; Dale et al., 2011; Fusaroli et al., 2014; Marwan et al., 2007), therefore, we limit our description to how the analysis relates to quantifying coordination between infant and adult vocalizations.

To obtain DCRPs, vocalization time series were divided into 1s bins. Each segment of either infant or adult vocalization was treated as occupying one time bin. This ensured that the interactivity estimated by the DCRPs was not affected by the durations of the segments, but only the timing between infant and adult vocalizations (Warlaumont et al., 2014).

Each DCRP returned the rate of co-occurrence of events across the two vocalization time series at 1s lags +/- 10s. Note that because overlapping segments between infant and adult vocalizations were excluded from all analyses, there are no lag- 0 recurrences reported here. DCRP height was computed by finding the total area under the DCRP profile between lag -10s and lag 10s. DCRP height measures the quantity of the infant-adult vocal interaction across a 10s sliding window. We estimated DCRP height for all three types of vocal interactions (infant speech-related and adult, infant non- speech-related and adult, and infant combined and adult) for each session.

Rate-based coordination

Vocalization rate was measured in terms of volubility, which was computed as the total amount of vocalization time in each recording sessions,divided by the duration of the recording session. Infant volubility measures were computed separately for speech-related vocalizations, non-speech-related vocalizations, and both types of infant vocalization. Adult vocalizations were not broken down by type. Volubility matching was measured in terms of the correlations between infant and adult volubilities across sessions and infants.

Cluster-based coordination

The hierarchically nested clustering of vocal onset events was estimated using Allan Factor analysis (Allan, 1966). Each time series of acoustic onsets was segmented into M adjacent and nonoverlapping windows of size T, then the number of events N_{j} was counted within each window indexed by $j = 1$ to M. The differences in counts between adjacent windows of a given size T were computed as $\mathrm{d}(T) = N_{\mathrm{j+1}}(T) - N_{\mathrm{j}}(T)$. The AF variance $\mathrm{A}(T)$ for a given timescale, T, is the mean value of the squared differences, normalized by two times the mean count of events per window (i.e., similar to coefficient of variation, but being constituted from differences between adjacent windows, whereas the typical coefficient of variation ignores temporal relations among elements),

$$A\left(T\right) = \frac{\left\langle d\left(T\right)^{2}\right\rangle}{2\left\langle N\left(T\right)\right\rangle .} \tag{4.1}$$

Poisson processes (i.e. random, independent events with exponentially distributed inter-event intervals) yield $\mathrm{A}(T) > 1$ 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, α the exponent of the scaling relation (Thurner et al., 1997), and $\alpha > 0$. This is a power law with positive exponent $\alpha$where αprovides a metric for the degree to which vocalization events are clustered across timescales. αcorresponds to the slope of the plots in panel D of Figure 1, which plots Allan Factor, A(T), vs. timescale, T, 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. AF slope does not necessarily reflect the degree of mature linguistic hierarchical structure although it does reflect the degree of hierarchical structure in the clustering of temporal events.

Ten timescales were used for all event time series. The time bins used were roughly the same across all recordings; there were small differences due to the dependency of the time binning algorithm on the total recording length. The average smallest timescale was ~10 s and the average largest timescale was ~88 min. Cluster-based coordination was measured by computing correlations between AF slopes for infant versus adult vocalizations. These correlations measure the extent to which the hierarchical clustering of infant vocalization bouts is similar to that of the adults in their environment across time.

4.4 Results

4.4.1 Volubility and Hierarchical Clustering Across Vocalization Types

First, we tested for differences in overall volubility across vocalization types. A one-way ANOVA with volubility as the dependent variable, vocalization type as the predictor variable, and infant as random intercept, indicated that volubility differed as a function of vocalization type, $F(3,2806)=456.02$, $p<.001$. A post-hoc Tukey test revealed that volubility for adult vocalizations (M=.06, SE=.002) was significantly higher than that for infant-combined vocalizations, i.e., non-speech-related and speech-related, (M=.05, SE=.0008), which was significantly higher than infant non-speech-related (M=.03, SE=.0006), which was significantly higher than infant speech-related (M=.02, SE=.0004), ps<.001.

Second, we tested for differences in hierarchical clustering across vocalization types. A(T) values and timescales were averaged across recordings and then A(T) was plotted as a function of T in Figure 2A. See Figure 2B for a scatterplot of each individual recording's values.

The linear trends in Figure 2 suggest that both infant and adult AF functions follow a power law. Flattening at the smallest timescales is expected to occur due to limitations in accuracy of the event onset labeling. To test against the null hypothesis that event time series are random (Pois-

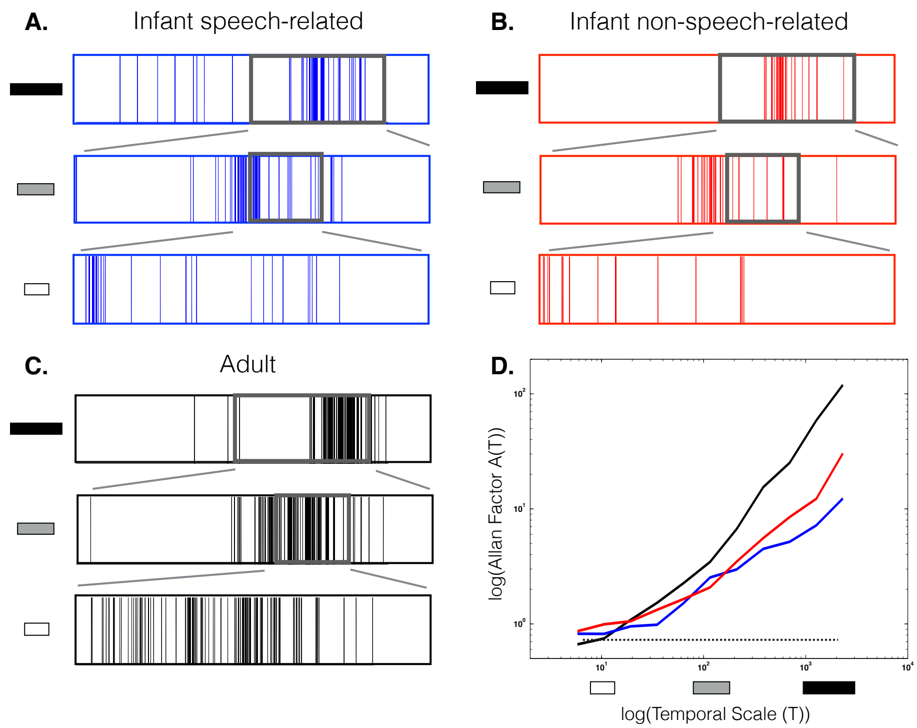

Figure 4.1: Schematic depiction of procedure of AF analysis at three timescales (∼7 minutes, ∼30 minutes, ∼60 minutes). (A-C) Vocalization events are counted within each timescale window. Each vertical line is an acoustic onset for one of the three vocalization types: (A) Infant speech-related, (B) Infant non-speech-related, and (C) Adult. The black, grey, and white rectangles indicate long (∼60 minutes), medium (∼30 minutes), and short timescales (∼7 minutes), respectively. Notice at each of the three timescales, there are clusters of onsets. AF variance is derived from computing the normalized squared difference of onset frequencies between adjacent time windows for the three timescales. AF variance is a measure of the departure from an equidistributed distribution of acoustic onsets. (D) The estimates of hierarchical clustering of vocalization types. The slope, α, of the $\log(AF)$ vs. $\log(T)$ curve estimates the scaling of AF variance across scales. The dotted line indicates a slope of 0 which is evidence for a random (Poisson process) vocalization event series. The other three curves have slopes closer to 1, indicating hierarchical clustering.

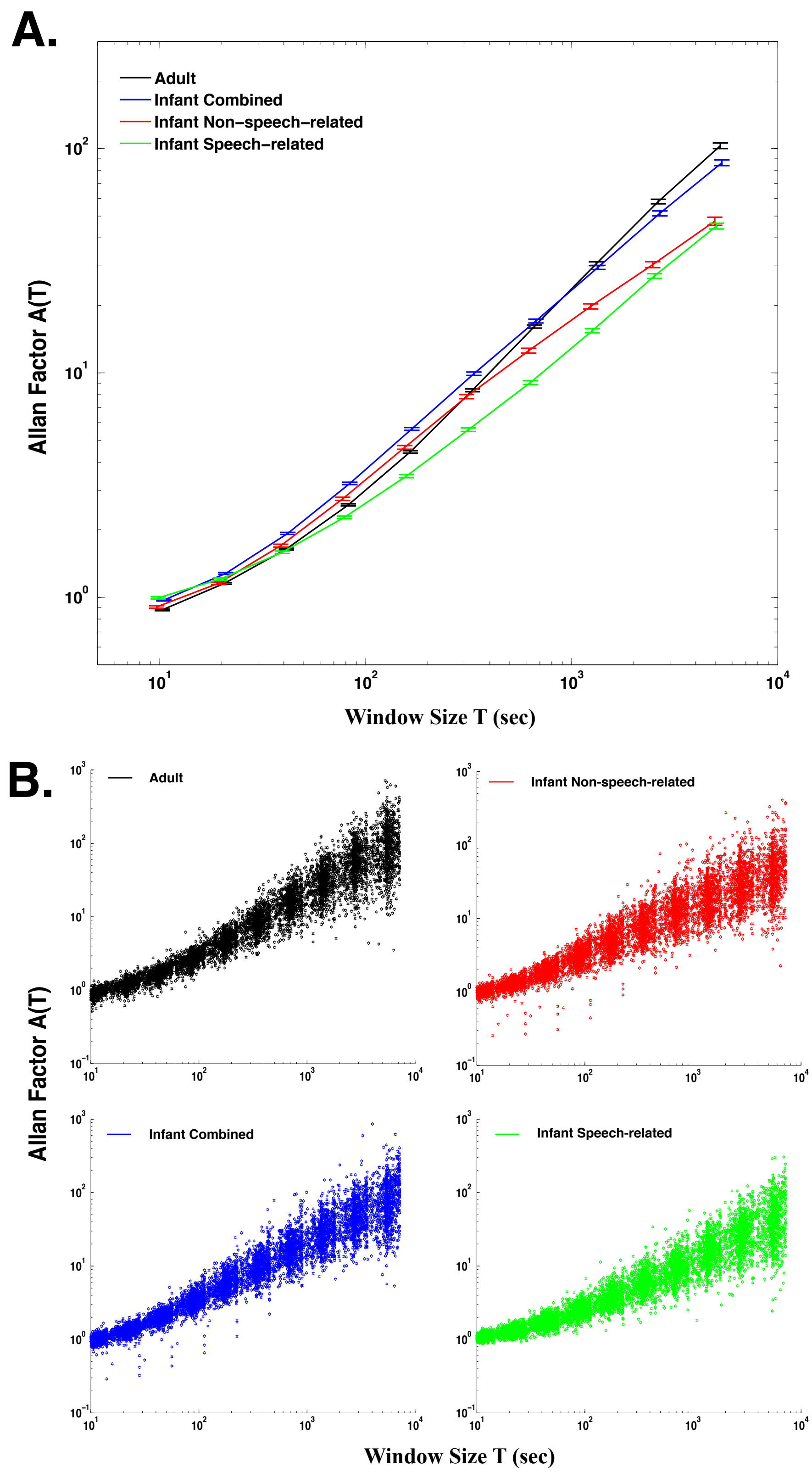

Figure 4.2: (A) Mean AF functions for adult and infant vocalizations, with standard error bars. (B) Scatterplot of each recording's A(T) values.

son distribution), we performed one-sample t-tests for AF slopes against a mean of 0. AF functions for all vocalization types were reliably greater than 0, ts(705)>147, ps<.001. Thus the positive linear trends in AF functions provide evidence that the onsets for all vocalization types were clustered across multiple timescales.

A one-way ANOVA with AF slope as the dependent variable, vocalization type as the predictor variable, and infant as random intercept indicated that the hierarchical clustering differed as a function of vocalization type, $F(3,2806)=413.17$, $p<.001$. A post-hoc Tukey test showed that AF slopes for the adult vocalizations (M=.76, SE=.004) were significantly steeper than for the infant-combined vocalizations (M=.71, SE=.004), which were in turn significantly steeper than infant non-speech-related (M=.62, SE=.004), which were significantly steeper than infant speech-related (M=.59, SE=.004), ps<.001. Shallower slopes indicate relatively less nesting of clusters in vocal onset events. (See Appendix for an additional power law analysis).

Finding the same pattern of effects on volubility and AF measures suggests that they may co-vary. Indeed, correlation analyses showed weak linear relationships between the two measures for infant speech-related (r[704]=.21, p<.001) and infant-combined (r[704]=.19, p<.001) vocalizations, and moderate relationships for adult (r[704]=.44, p<.001) and infant non-speech-related (r[704]=.41, p<.001) vocalizations. Volubility and AF measures appear to reflect one or more common sources of variation, but also exhibit unique effects, as the following analyses show.

To determine whether there was change in volubility and hierarchical clustering over the first year of the infants' lives, we regressed AF slope and volubility on infant age, performing separate analyses for the three types of infant vocalizations and the adult vocalizations. To determine unique effects on each dependent measure, all subsequent analyses were conducted by first computing the correlation between volubility and. hierarchical clustering estimates, then obtaining the residual values of either volubility or hierarchical clustering after factoring out their correlation. We then tested for a relationship between the residual values and other variables of interest. For example, if we were interested in the relationship between hierarchical clustering of infant-combined vocalizations and age of infant, we would first compute the residual values of hierarchical clustering after factoring out the (linear) relationship between hierarchical clustering and volubility of infant-combined vocalizations. We then tested if the residual (unique variance of hierarchical clustering) correlated with age of infant using a first-order correlation, r_{residual}. To control for infant-level variance, we computed the residuals using linear mixed effects models with infants as random intercepts (Baayen, Davidson, & Bates, 2008). We also present the results of correlation analyses without other variables factored out, to show whether the directions of any effects changed as a result of residualization. Although we present both the first-order correlations (r) and the correlation coefficients from the residuals analyses (r_{residual}), we interpret all results using the magnitude, direction, and statistical significance of the r_{residual} values.

Table 4.1: Results of first order correlations and residual analyses predicting infant age.

All Infant Vocalizations	r	r_{residual}
Volubility	.002	.01
AF	-.10**	-.12**
Speech-related		
Volubility	.20***	.26***
AF	.05**	-.03
Non-speech-related		
Volubility	-.15***	-.16***
AF	-.24***	-.25***
Adult		
Volubility	-.24***	-.19***
AF	-.28***	-.18***

Note. #p<.1, *p≤.05, **p≤.01, ***p≤.001. For all analyses, degrees of freedom = 704. AF = Allan Factor estimate.

Table 1 shows how AF and volubility vary as a function of age. Volubility increased with infant age for infant speech-related vocalizations, and decreased with infant age for infant non-speech-related vocalizations and adult vocalizations. No change in volubility was observed for infant vocalizations when both speech-related and non-speech-related vocalizations were included. AF slopes decreased for infant vocalizations overall as well as for non-speech-related vocalizations, but did not change with age for infant speech-related vocalizations. AF slopes also decreased with age for adult vocalizations. We discuss the implications of this below when presenting results on the relation between infant and adult AF slopes.

4.4.2 Do coincidence-based, rate-based, and cluster-based coordination patterns vary depending on the type of vocalization produced by the infant?

The primary goal of the current study is to investigate the different vocal coordination patterns of infant and adult vocalizations. For each of the three coordination patterns, we (1) assessed whether the coordination pattern existed beyond baseline controls, (2) whether such coordination patterns still held after controlling for the other coordination patterns, and (3) if the degree of the coordination differed as a function of the vocalization type produced by the infant.

To measure coincidence-based vocal coordination between infants and adults we used DCRP height, derived from CRQA. Higher DCRP heights suggest more coincidence-based vocal coordination. The first step was to set up a baseline measure to compare against empirical pairings of infant and adult vocalization series. Our baseline measure consisted of shuffling the empirical infant and adult time series then submitting them to CRQA to get baseline DCRP height. We chose this baseline measure because it preserves the number of vocalizations and it keeps the shuffled time series the same length as the original time series. We obtained the DCRP height and baseline DCRP height for all three vocalization types. A one-way ANOVA with infant as random intercept indicated that DCRP height for the original time series was on average higher than shuffled DCRP height across all vocal interaction types, $F(1,4220)=221.68$, $p<.001$. Because shuffled DCRP height differed as a function of vocal interaction type, we normalized the original DCRP height by subtracting the corresponding shuffled DCRP height from the original DCRP height for each vocal interaction type. A one-way ANOVA with normalized DCRP height as the dependent variable, vocal interaction type as the predictor variable, and infant as random intercept indicated that the degree of coincidence-based coordination differed as a function of vocal interaction type, $F(2,2101)=74.81$, $p<.001$. A post-hoc Tukey test showed that normalized DCRP heights for the infant-combined and adult vocalizations (M=.001, SE=.00009) were significantly taller than those for the infant speech-related and adult vocalizations (M=.0009, SE=.00005), which were in turn significantly taller than infant non-speech-related and adult vocalizations (M=.0002, SE=.00004), ps<.001. The same patterns of differences were found when using non-normalized DRCP heights. These results suggest that there was coincidence-based coordination above and beyond a random baseline. Furthermore, coincidence-based coordination was stronger for speech-related relative to non-speech-related interactions. See Figure 3 for DCRPs for the three vocalization types.

To determine the degree of rate-based and cluster-based coordination between infant and adult vocalizations we correlated volubility and AF slopes measured for adult vocalizations with those for each of the three corresponding infant vocalization types. Correlations were computed between raw infant and adult measures as well as between residuals of the infant and adult measures after taking out any correlation with age of infant and AF slope, volubility, or DCRP height (whichever two were not the focus of a given comparison). For example, to assess cluster-based coordination, infant and adult AF slopes were each residualized against

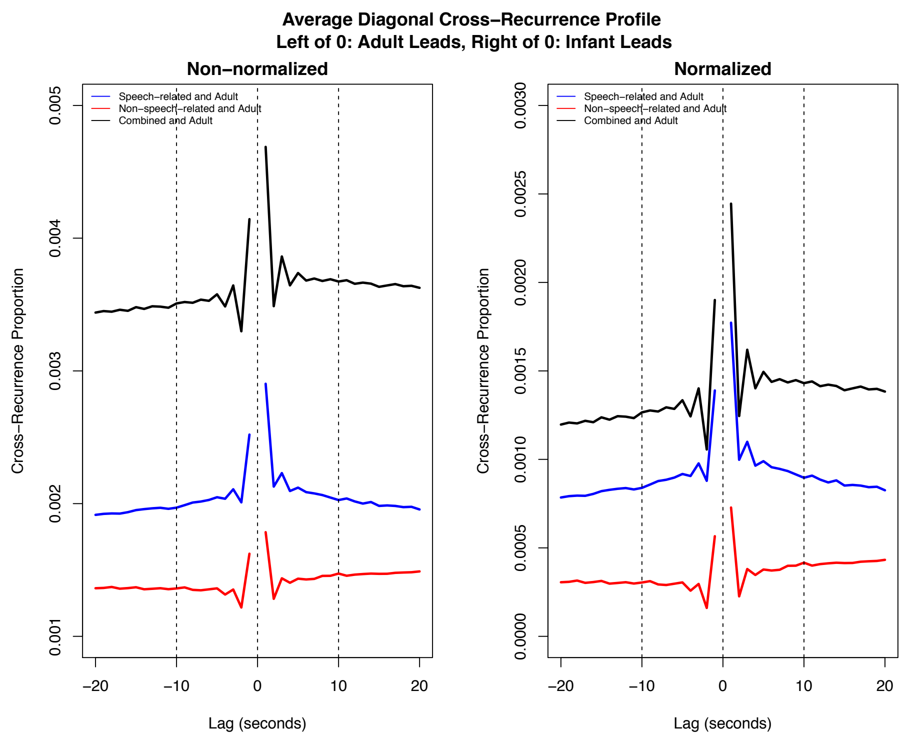

Figure 4.3: Diagonal cross-recurrence profile (DCRP) averaged across all vocalization types. (Left) Average DCRPs are before normalization. (Right) Average DCRPs normalized for shuffled DCRPs.

speech-related volubility of the same speaker type, speech-related DCRP height, and age of infant. As before, to control for infant-level variance, we computed the residuals using linear mixed effects models with infant as random intercepts. See Table 2 and Figure 4 for results.

Table 4.2: Results of first order correlations (r) and residual correlations (r_{residual}) predicting matching of infant vocalization properties with adult volubility and adult AF slope estimates.

Rate-based Vocal Coordination	r	r_{residual}
All Infant Vocalizations	.26***	.10***
Speech-related	.21***	.13***
Non-speech-related	.23***	.06
Cluster-based Vocal Coordination		
All Infant Vocalizations	.15***	.20***
Speech-related	.14***	.25***
Non-speech-related	.14***	.04

Note. $^{\#}$p<.1, *p≤.05, **p≤.01, ***p≤.001. For all analyses, degrees of freedom = 704.

For rate-based coordination, Infant-combined and infant speech-related vocalization types reliably matched the volubility pattern of adult vocalizations. Using the Fisher r-to-z transformation to test for differences between correlation strength, infant speech-related volubility matching was marginally stronger than matching between infant non-speech-related vocalization, $z = 1.75$, $p = .08$. For cluster-based coordination, infant combined and infant speech-related vocalization types reliably matched the structure found in adult vocalizations. Cluster-based vocal coordination between adult vocalizations and infant speech-related vocalizations was significantly stronger than matching between adult vocalizations and infant non-speech-related vocalizations, $z = 4.25$, $p < .001$.

4.4.3 Are adults or infants primarily driving these vocal coordination patterns, and does this change with age?

In the previous section, we observed that different measures of vocal coordination were not statistically reducible to each other. Thus these measures appear to provide unique information about the relationships between infant and adult vocalization properties. In this section, we explore the question of what information the different vocalization measures provide about whether it is infants or adults who are the primary drivers of vocal coordination during the first two years of life.

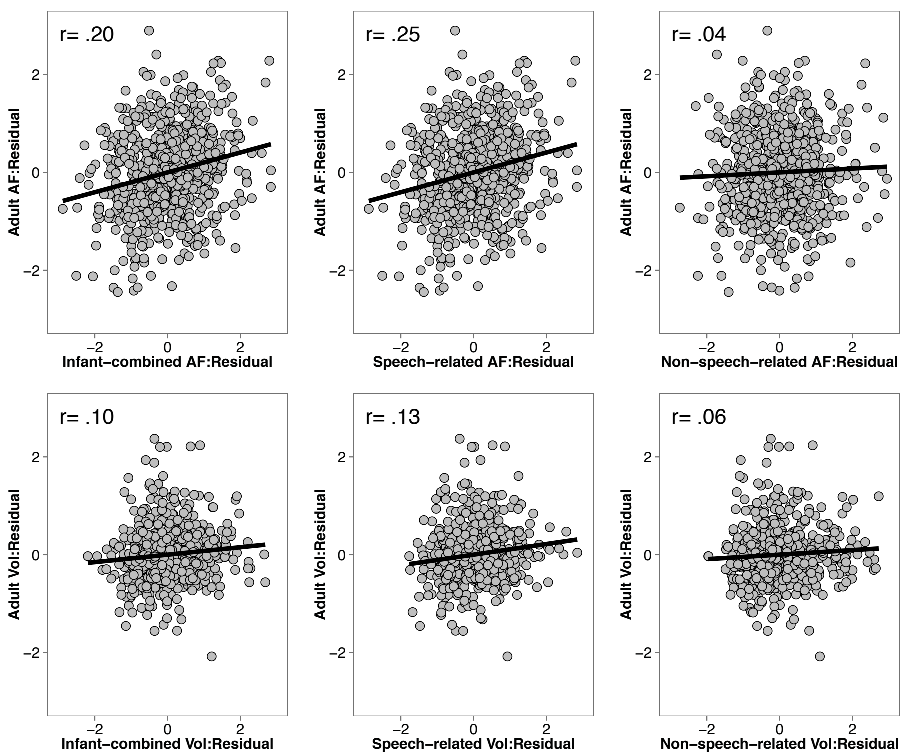

Figure 4.4: (Top row) Cluster-based vocal coordination results for Adult and (left to right) Infant-combined, Infant-speech-related, and Infant-non-speech-related. (Bottom row) Rate-based vocal coordination results. All variables are standardized. Each circle represents an individual recording.

For coincidence-based coordination, we can measure leader-follower patterns in vocalizations. We computed a leader-follower ratio from the original DCRP for each coincidence-based coordination by taking the ratio of the sum of the right side (infant leading side) to the sum of the left side (adult leading side) of the ±10s DCRP profile (Warlaumont et al., 2014). A leader-follower ratio greater than 1.0 indicates that the infant led the adult whereas a ratio less than 1.0 indicates the adult led the infant.

A one-way ANOVA with leader-follower ratio as the dependent variable, vocal interaction type as the predictor variable, and infant as random intercept indicated that infant leading differed as a function of vocalization type, $F(2{,}2101)=14.85$, $p<.001$. A post-hoc Tukey test showed that leader-follower ratios for the infant-combined and adult vocalizations ($M=1.049$, $SE=.002$) were higher than the ratios for infant non-speech-related and adult vocalizations ($M=1.041$, $SE=.002$, $p=.006$), and infant speech-related and adult vocalizations ($M=1.035$ $SE=.002$, $p<.001$). Leader-follower ratios for infant non-speech-related and adult vocalizations were higher relative to ratios for infant speech-related and adult vocalizations, $p=.048$.

To determine whether leader-follower ratios changed across infant age, we tested for correlations between ratios for each vocalization type and infant age. We observed no reliable association between infant speech-related ($r[704]=-.05$, $p=.19$) or infant-combined ($r[704]=-.05$, $p=.19$) leader-follower ratios and age. We did observe a reliable negative association between infant non-speech-related leader-follower ratios and age ($r[704]=-.08$, $p=.04$), suggesting that as infants grew older, there was a decrease in the tendency for infant non-speech-related vocalizations to precede adult vocalizations rather than vice versa.

For volubility and hierarchical clustering, we computed absolute similarity scores and then tested for correlations between the difference scores and infant age. For the difference score (SS), we computed an absolute similarity score by subtracting infant vocalization property (AF or Volubility) from the adult vocalization property, taking the absolute value, and subtracting the value from 1, e.g.,

$$SS_{AF} = 1 - abs(AdultAFSlope - InfantAFSlope.) \tag{4.2}$$

A similarity score of 1.0 suggests the vocalization properties across infant and adult were identical. A positive correlation between SS and age indicates greater matching between infant and adult on that characteristic as age increased. Figure 5 provides a graphical depiction of these results.

Adults and infants showed statistically significant increases in coincidence-based vocal coordination for all infant vocalization types (speech-related: $r[704]=.27$, $p<.001$; non-speech-related: $r[704]=.21$, $p<.001$, all: $r[704]=.21$,

p<.001) as well as in cluster-based vocal coordination for infant speech-related vocalizations (r[704]=.18, p<.001) but not in cluster-based vocal coordination for infant all vocalizations (r([704]=.04, p=.25), or infant non-speech-related vocalizations, r(704)=.05, p=.18.

Using the Fisher r-to-z transformation to test for differences between correlation strength, we observed stronger convergence for speech-related vocalizations relative to non-speech-related vocalizations for both volubility (z=9.04, p<.001) and hierarchical clustering, $z = 7.31$ p<.001.

For infant speech-related hierarchical clustering, combining the observation that infants and adults converge with age with the result that infant hierarchical clustering does not change with age and the result that adult hierarchical clustering decreases with age, we can infer that the adult vocalization environment is adapting its hierarchical clustering to that of the infant over the course of the first two years of life. Because infant speech-related volubility increases whereas adult volubility decreases over infant age, the results from the difference score analyses suggest bidirectional convergence: Both infants and adults adjust volubility rates towards each other over infant age.

4.4.4 Do the different coordination measures have unique developmental trends?

In the previous sections, we established that the three vocal coordination patterns are not reducible to each other and provide different perspectives on the interpersonal dynamics of infant-adult vocal coordination. In this final section, we investigate whether the various coordination patterns are independently associated with infant age.

In addition to the three vocal coordination patterns that have been the foci of this study, for this section we also included a conversational turn taking measure computed by the LENA system. The conversational turn taking measure computed by LENA is frequently used in the literature (Caskey, Stephens, Tucker, & Vohr, 2011; Gilkerson & Richards, 2008; Gilkerson, Richards, & Topping, 2015; Greenwood et al., 2011; Suskind et al., 2015; Warren et al., 2010) and is therefore an important measure to include when assessing independent associations with infant age. A conversational turn is identified when a sequence of speech-related sound segments from an adult then an infant, or vice versa, occurs within 5s without an intervening non-speech-related segment or speech-related segment from another adult or infant. Conversational turn count can be considered a measure of infant-adult interaction (Warren et al., 2010). Because recording sessions in our sample greatly varied in length, we computed turn taking rate by dividing conversational turn count by the length of the recording session.

Because the turn taking rate is computed using only speech-related seg-

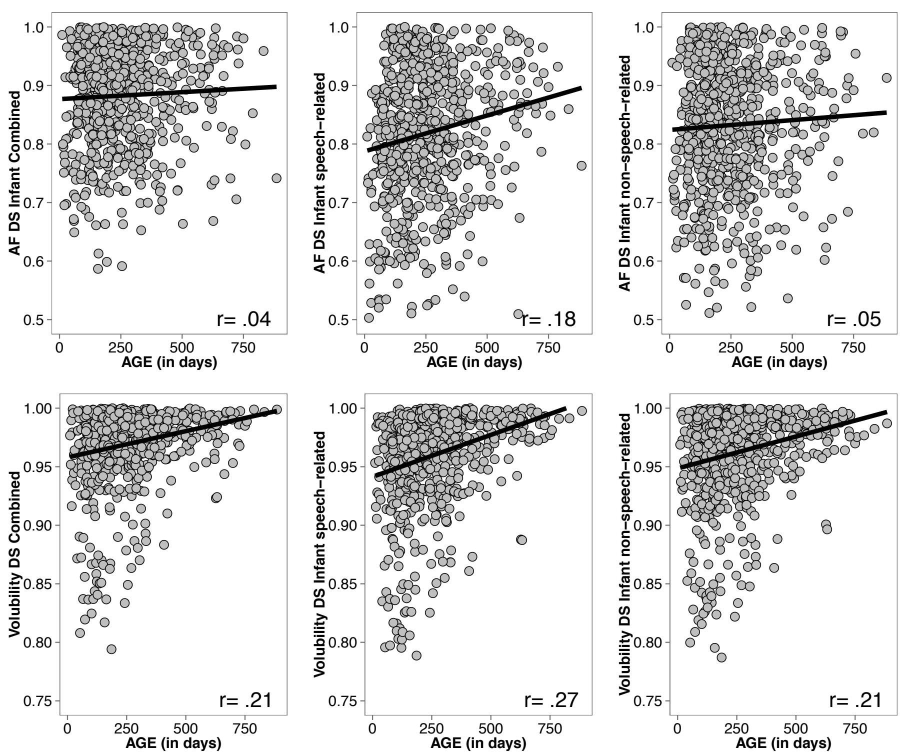

Figure 4.5: (Top row) Difference Score (DS) results for infant age and (left to right) Infant-combined hierarchical clustering estimates, speech-related hierarchical clustering estimates, and non-speech-related hierarchical clustering estimates. (Bottom row) DS results for infant age and (left to right) infant-combined volubility, speech-related volubility, and non-speech-related volubility. Note. AF and Volubility DS axes have different ranges.

ments, we limited our analyses in this section to speech-related coordination patterns. Table 3 reports first-order correlations and also correlations with residualized variables. Coincidence-based coordination (r_{residual}=.07, p=.05), rate-based coordination (r_{residual} =.31, p<.001), and conversational turn-taking rate (r_{residual} =.15, p<.005) were all independently positively associated with infant age. Cluster-based coordination was not independently associated with infant age, r_{residual} =.02, p=.61.

Table 4.3: Results of first order correlations (r) and residual correlations (r_{residual}) of coordination patterns and infant age.

Coordination Pattern	r	r_{residual}
Coincidence-based	<.001	.07*
Rate-based	.27***	.31***
Cluster-based	.18**	.02
Turn-taking rate	-.03	.15***

Note. #p<.1, *p≤.05, **p≤.01, ***p≤.001. For all analyses, degrees of freedom = 704. Rate-based and Cluster-based coordination reflect difference scores.

4.5 Discussion

This study examined coordination patterns that arise from different measures of infant and adult vocalizations. We aimed to answer three specific questions: (1) Do coincidence-based, rate-based, and cluster-based coordination patterns vary depending on the vocalization type produced by the infant? (2) Do different coordination patterns provide unique information about the interpersonal dynamics of vocal interaction? (3) How do the various coordination patterns relate to infant age?

We observed that all three coordination patterns displayed higher rates of coordination for infant speech-related vocalizations relative to infant non-speech-related vocalizations. These results point to a difference in coordination as a function of speech- relatedness, and could perhaps be due to speech-related vocalization holding more social value to caregivers. Properties derived from the coordination patterns provided new insights into unidirectional and bidirectional adaptation between infants and their caregivers. Finally, we observed unique trajectories between the coordination patterns and infant age.

4.5.1 Hierarchical vocalization patterns and volubility

To answer the three research questions provided at the outset of this paper, estimations of vocalization properties such as hierarchical clustering and volubility were required. An important finding from this study was that the onsets of infant vocalization bouts have hierarchical structure at timescales ranging from seconds to hours. This result expands upon previous work using subjective ratings to assess hierarchical structure or phrasing of infant vocal productions at shorter timescales (Lynch et al., 1995) and also converge with evidence of hierarchical structure in speech based on other algorithms evaluating other vocalization patterns (Abney, Warlaumont, Haussman, Ross, & Wallot, 2014; Abney, Paxton, et al., 2014; Luque et al., 2015). Lynch et al. identified the hierarchical organization of syllables, utterances, and prelinguistic phrases, and identified hierarchical structure spanning the typical duration of syllables (<500 ms) to less than several seconds in duration. Because of the temporal resolution of the automated vocalization segmentation used in our study, the shortest timescale included in our estimate of hierarchical clustering was approximately 10s. The hierarchical structure we identified spanned from $\sim$10s to $\sim$1.5hrs. Therefore, the hierarchical clustering observed in the present study is at the level of bouts of vocalization and does not reflect the structure within utterances. Future work is required to better understand the hierarchical structure of infant vocalizations at shorter timescales, e.g., spanning milliseconds to seconds. These results also suggest that infant prelinguistic vocalizations are not equidistributed and are power-law distributed. Follow-up analyses (not reported here) demonstrated that the inter-event intervals of the vocalization events were power-law distributed with a slope approximating -2. Our results therefore provide evidence for fractal properties of prelinguistic vocalizations.

Evidence for hierarchical clustering of vocalizations was found at even the youngest session, recorded from an infant who was 11-days-old. Although estimates of hierarchical clustering for infant speech-related vocalizations were not observed to change with age, we observed a reliable decrease of hierarchical clustering (more random) for infant non-speech-related vocalizations. The results presented here suggest that infant vocalization bouts exhibit non-random temporal patterning from shortly after birth and that, for speech-related vocalizations, this hierarchical nature of vocalization bouts is fairly stable across the period of prelinguistic and early linguistic development.

We also investigated patterns of infant volubility. Previous work has suggested that by about 3–5 months of age, infants learn that vocalizations have social value, with more communicative types of vocalizations influencing parental engagement (Goldstein et al., 2009). Previous work has also found that adult responsiveness to infant vocalizations increases during the second year (L. Bloom, Margulis, Tinker, & Fujita, 1996). In the present study, volubility for infant speech-like vocalizations increased with infant age, replicating prior findings that also used the LENA system (Greenwood et al., 2011) and strengthening the idea that, over time, infants learn that vocalizations hold social value and serve a communicative function. We also found that volubility for infant non-speech-related vocalizations decreased with infant age (similar to Warlaumont et al., 2014).

It is important to point out a few possible limitations to the observed results. It is always possible that the increases in volubility are influenced by decreasing sleeping time relative to neonates. Although this is a possibility, naps are a component of an infant's daily routine and among the many factors of the complex interaction between infant vocalization bouts and adult vocalization bouts. Additionally, the ability of the LENA system to discern infant vocalizations may improve with age. Therefore, it is possible that changes in volubility across age are at least partially due to differences in the ability of the LENA system to discern between infant vocalizations across age. Future work combining automatic and manual coding procedures is important to establish the reliability of increased volubility across age.

We found that changes in hierarchical clustering and in volubility across age held even when other variables were factored out through residualization. These results, combined with the different developmental patterns observed for volubility vs. hierarchical clustering, suggest that volubility and hierarchical clustering provide at least partially independent information about infant prespeech and early speech development. The estimation of hierarchical clustering of vocalizations may provide additional measures that can help predict later infant behaviors and abilities. For example, the hierarchical clustering of infant behavior may reflect the daily routines of a family and/or daycare environment, and the predictability of these routines may be reflected in the consistency of AF slopes across recordings. Future work is required to test whether or not hierarchical clustering is a vocalization property with predictive value for important developmental outcomes.

4.5.2 Vocal coordination patterns vary by vocalization type and provide unique information based on level of description

We introduced a typology of coordination patterns that spans across levels of description and time scale: coincidence-based, rate-based, and cluster-based vocal coordination. Using CRQA, we observed that coincidence-based coordination was greater than a random baseline based on shuffled time series. One potential issue with the data collection technique used in the current study is that we are not directly aware of specific bouts of interaction relative to incidental vocalizations made by infants and adults in the infants' auditory environments. Showing that empirical DRCP heights were greater than surrogate-based DRCP heights provides evidence for the non-incidental, vocal interaction of infants and adults in close proximity to the infant.

Across the different vocal coordination patterns, we found that coordination patterns based on infant speech-related vocalizations were stronger and more frequent relative to coordination patterns based on infant non-speech-related vocalizations. These results point to the sensitivity of the coordination patterns based on child vocalization type.

4.5.3 Different coordination patterns provide unique information about the dynamics of vocal interaction

For coincidence-based vocal coordination, we computed leader-follower dynamics across vocalization type and across temporal lag. We found that within a 10s window, infant vocalizations precede adult vocalizations and more so for non-speech- related vocalizations. Rate-based and cluster-based vocal coordination patterns offered a different perspective on leading and following in vocal dynamics. Focusing on rate-based patterns, we found bidirectional convergence of volubility across infant age: infants and adults both adjusted volubility rates towards each other across age. Focusing on cluster- based vocal coordination, we found that adults adapted the hierarchical clustering of their vocalizations to that of their infants' vocalizations as infant age increased.

Also studying daylong home audio recordings. Ko, Seidl, Cristia, Reimchen, & Soderstrom (2015) investigated the relationship between acoustic properties of mother and infant/toddler vocalizations. Ko et al. observed that mothers and infants/toddlers converged across various vocalization properties such as pitch. Specifically, mothers adapted their speech to the infant/toddler more than vice versa. The results of the current paper extend what Ko et al. observed by pointing to another vocalization property, hierarchical clustering, that shows similar convergence patterns. Notably, there was adult-to-infant convergence of both hierarchical clustering and volubility. Our results diverge from Ko et al. in the timescales of convergence: Ko et al. found convergence of pitch at the local level of conversational exchange whereas the results in the current study found convergence of hierarchical clustering and volubility across the entire span of daylong recording session, e.g., $\geq$6hrs.

The observation that adults adapted the hierarchical clustering of their vocalizations to that of their infants' vocalizations adds additional support to the fine- tuning hypothesis (Snow, 1989, 1995),suggesting that adults adapt the complexity of their child-directed language in response to properties of child-produced language. Most of the support for the fine-tuning hypothesis focused on measures of linguistic complexity (Kunert, Fernández, & Zuidema, 2011; Snow, 1995; Sokolov, 1993). Our results support the fine-tuning hypothesis, but use a metric focused on the hierarchical organization (a hallmark of 'complex systems') of vocal clustering instead of linguistic complexity. Future work testing the fine-tuning hypothesis should consider multiple measures of 'complexity' spanning various levels of linguistic and vocal alignment.

4.5.4 Coordination patterns and infant age

Since Bateson (1975) and Stern et al., (1975) first proposed that an important property of interpersonal exchange and communicative function was the development of turn taking dynamics, several studies have illuminated developmental patterns of vocal interaction (Caskey et al., 2011; Harder et al., 2015; Hilbrink, Gattis, & Levinson, 2015). These studies provide important information about the timing of turn taking (e.g., Hilbrink et al., 2015) or the transition from covocalizations to turn taking across development (e.g., Harder et al., 2015; Rutter & Durkin, 1987). But turn taking is only one type of vocal coordination. Our investigation of multiple vocal coordination patterns across development adds to prior research by showing the relationships between vocal coordination patterns focusing on different levels of analysis, and infant age. We found that different coordination patterns had different associations with infant age. Rate-based vocal coordination had the strongest association with infant age: speech-related rate-based vocal coordination increased with infant age. Turn-taking rate and coincidence-based coordination both increased with infant age as well. When controlling for all other coordination patterns, cluster-based coordination was not associated with infant age. Although cluster-based speech-related vocal coordination did not change significantly with increasing infant age once other coordination patterns were controlled for, cluster-based coordination may nevertheless reflect an aspect of coordination between infant and caregivers that has developmental significance, e.g., by facilitating information transfer between infant and caregiver across the first year (see paragraphs below).

We found that infants' vocal timing became more similar to their caregivers' vocal timing across the first two years of life. In other words, within a 10-second temporal window, infant and caregiver vocalizations occurred more frequently across infant age. This finding, in conjunction with the results of increased turn taking rate and increased rate-based matching across age suggests a dynamic trajectory of vocal development. Throughout the first few years of life, infant and caregiver vocalizations become more temporally coordinated (coincidence-based vocal coordination), vocalize at similar rates across the day (rate-based vocal coordination) and increase the rate of structured turn taking patterns (turn-taking rate).

4.5.5 Future directions

An important potential application of infant-adult vocal coordination patterns is to the study of language development and atypical development. Jaffe et al.'s (2001) contribution is an example of the utility of using coordination patterns to predict developmental outcomes. Future work should incorporate a pluralistic approach to coordination patterns to determine the predictive value of different coordination patterns for important developmental outcomes. To that end, it is important to understand what information different coordination patterns provide.

Coincidence-based vocal coordination provides information about the similarities and differences in vocal timing. Rate-based vocal coordination provides information about the similarities and differences in overall volubility rates across a recording session. Cluster-based vocal coordination provides information about the similarities and differences in the production of hierarchical clustering across a recording session.

Although all three coordination patterns provide important information about vocal interaction, cluster-based vocal coordination is motivated by a theory in statistical mechanics investigating the outcomes of interacting complex networks. More specifically, work in statistical mechanics has showed that when two complex systems interact, information transfer between them is enhanced and may even become optimal when their multiscale dynamics are matched (West et al., 2008), a term called complexity matching. Previous research studying adult conversations has shown that the degree of cluster-based vocal coordination or complexity matching differed depending on specific conservational contexts (Abney, Paxton, et al., 2014). Perhaps a function of cluster-based vocal coordination is increased communication? Indeed, the question of function for any coordination pattern or collection of coordination patterns should be the focus of future research.

This information transfer hypothesis requires much more empirical attention before any substantive conclusions can be made. For example, recent work on infant language development has utilized the LENA system along with various standardized measures of language and communication development (e.g., MacArthur-Bates, Communicative Development Inventory; Fenson et al., 2007) to investigate language learning in naturalistic environments (Ramírez-Esparza et al., 2014; Weisleder & Fernald, 2013). Future studies should investigate the role of the production and convergence of specific vocalization properties like volubility and hierarchical clustering on vocabulary or other aspects of language development (Northrup & Iverson, 2015).

4.6 Conclusion

Our results support the proposal that various vocal coordination patterns spanning multiple levels of description provide unique information about infant-adult vocal interactions. We found increased coincidence-based, rate-based, and cluster-based vocal coordination for infant speech-related vocalizations relative to non-speech-related vocalizations. We also found different infant-adult convergence patterns depending on the measure used. For instance, leader-follower dynamics derived from coincidence-based coordination measurements suggest that infants lead vocal exchanges whereas adults adapt their hierarchical clustering to that of the infant over time. Finally, we found divergent associations between infant age and the various vocal coordination patterns. In particular, higher degrees of speech-related coincidence-based, rate-based, and conversational turn taking were independently associated with increased rates of turn taking. Future work should address the question of how the various coordination patterns relate to the different contexts and event types the infant experiences over the course of the day and should attempt to discover the unique functions the different coordination patterns serve (if any). Future work should focus on utilizing multiple vocal coordination patterns in combination to test whether multiple levels of description increase the predictive value for identifying important developmental milestones or diagnosing various clinical disorders.

Chapter 5

Discussion

5.1 Introduction

Research on human communication is a case study on interdisciplinarity, incorporating diverse fields of science, engineering, mathematics, and the humanities. At the outset of this dissertation, I introduced the Complexity Matching hypothesis for human communication. The Complexity Matching hypothesis for human communication predicts that when the hierarchical structure of communicative patterns match between two people, information transmission is enhanced.

Along with investigating the Complexity Matching hypothesis for human communication, this dissertation addressed the question of whether there is function to the behavioral and coordination patterns during human conversation and interaction. Chapter 2 introduced the notion of Complexity Matching as a potential function of information transmission in adult conversations. Chapter 3 presented a preliminary test of the hypothesis of information transmission. Chapter 4 investigated the development of the production and convergence of hierarchical patterns in vocal production and interaction. Below, I summarize the important observations and contributions from each of these chapters.

5.2 Production and convergence of hierarchical structure

Across the three studies presented in this dissertation, I focused on the production and convergence of hierarchical structure in vocalization and movement behaviors. I observed that both speech and movement are hierarchical and nested across multiple temporal scales.

Multiscale clustering of human behavior reflects hierarchical nested patterns across temporal scales. For speech, this is related to the notion of hierarchical structure in language. From phonemes (Pardo, 2006) and syntax (Bock, 1986) to language style (Niederhoffer & Pennebaker, 2002; Ireland et al., 2011) and turn taking timing (Manson et al., 2013), there are many examples of the matching, alignment, or convergence of specific levels of linguistic representation between interlocutors. In Chapter 2, I observed that the hierarchical structure of speech differs as a function of conversational context. Specifically, argumentative conversations have more correlated clustering of speech onsets relative to affiliative conversations. Follow-up analyses showed that argumentative conversations have stronger anti-correlated speech activity, suggesting that there were stricter, more structured, turn-taking dynamics, relative to affiliative conversations. Therefore, having higher correlated clustering of onsets across multiple temporal scales might equate to more structured patterns of speech onsets.

Relative to speech, the observation that there are hierarchical nested patterns of overall movement behavior is less understood. There is a large collection of research displaying $1/f$ scaling and long-range correlations in motor performance (Stephen et al., 2010; Stephen & Hajnal, 2011; Palatinus, Kelty-Stephen, Kinsella-Shaw, Carello, & Turvey, 2014; Riley & Turvey, 2002; Wijnants, Bosman, Hasselman, Cox, & Van Orden, 2009; Stergiou & Decker, 2011). The observation of multiscale clustering of movement onsets adds to this collection of literature showing hierarchical structure in specific movement patterns. However, the observation of multiscale clustering in movement onsets requires more description with potential insights into patterns of variability or bouts of activity that are nested across multiple temporal scales.

In Chapter 3, I observed multiscale clustering of movement onsets during a dyadic problem-solving task. During the task, there are specific behaviors that the dyad members would have to perform on a regular basis such as reaching for objects, stabilizing an object, and gesturing an idea. All of these bouts of movements likely recur throughout the construction phase. Going even further, it is likely combinations of these bouts of movement occur and recur as well. For example, we could imagine a sequence of movement bouts that recur such as (1) Person A gestures an idea, (2) Person A picks up an object, and (3) then Person A stabilizes the object while Person B connects two objects together. Future work is required to better understand the hierarchical nested patterns of movement behavior across interaction contexts.

The convergence of hierarchical patterns of speech and movement is consistent with the notion of the functional hypothesis of complexity matching. Indeed, this is a key insight of the Complexity Matching hypothesis. In Chapter 2, I observed that the degree of matching of hierarchical patterns of speech differed as a function of conversational context. In Chapter 3, I observed that the degree of matching of multiscale clustering patterns of vocalizations and movement predicted performance on a dyadic problem-solving task. Crucially, the observations in Chapter 3 provided preliminary evidence for this dissertation's hypothesis that there is function to the structure of multiscale communicative behaviors.

Additional attention must be given to the possibility that complexity matching might not be independent from local coordination patterns of the two systems. For a functional hypothesis of converging complex phenomena, is it the global coordination patterns of the two complex systems, the local coordination patterns, or a mixture of coordination patterns that explain information transmission? In Chapter 5, we found that multiple coordination patterns provide unique information about vocal development and interaction.

Questioning the potential mutual influence of local and global coordination patterns requires additional inquiry into the coupling medium of the two systems. In the famous case of Huygen's coupled pendulum clocks, two pendulum clocks become synchronized when they were mechanically coupled by being mounted on the same base structure. It is the motion of the base structure that determines the rate of synchronization (see Pantaleone, 2002, for an overview). It has been shown that visual (M. J. Richardson et al., 2005), auditory-speech (Shockley et al., 2007), and haptic (Marmelat & Delignières, 2012) coupling between humans can lead to local coordination patterns like synchronization. However, what is the coupling medium for global coordination? Clues to this answer can be seen in work by Marmelat and Delignières (2012). They found that the visual and haptic coupling between two people led to global coordination patterns that were independent of any observed local coordination patterns. However, what does this coupling reduce to? It is possible that the coupling medium between two people during an interaction reduces down to explanations at the level of neuronal and large-scale synchrony of brain regions. Indeed, recent work has provided evidence for brain-to-brain coupling during various interaction tasks (Stephens, Silbert, & Hasson, 2010; Dikker, Silbert, Hasson, & Zevin, 2014; Hasson, Ghazanfar, Galantucci, Garrod, & Keysers, 2012).

5.3 Quantification of multiscale clustering of communicative behaviors

In this dissertation, novel methods were developed to estimate the multiscale clustering, and by extension, the hierarchical structure, of communicative behaviors. The Allan Factor (Allan, 1966) was adapted to estimate the clustering of event onsets of communicative behaviors like vocalizations and body movements. This is a novel methodological contribution to many diverse literature areas.

Assessing the hierarchical complexity of language has been a considerable research program since the late 1960s (Miller & Chomsky, 1963; S. C. Johnson, 1967; Levelt, 1969; Suci, Ammon, & Gamlin, 1967). However, most of the techniques used to assess hierarchical structure relied primarily on human raters or breaking down various sentence structure types. The utilization of the Allan Factor analysis to estimate multiscale clustering of events provides a new method for assessing hierarchical structure in vocalizations and other behaviors. It is important to note major differences between previous methods and the AF analysis used herein.

First, the AF analysis requires a binary spike train, where '1' denotes a specific event. The user defines the conditions required in the data for an event to occur, and is therefore an important set of free parameters that need to be justified and communicated effectively for transparency. The utilization of a binary spike train of events diverges from the previous methods that were more directly mapped onto the linguistic properties of sentence structure. Second, the AF analysis can be used to estimate the hierarchical structure of event series without the need of human raters, which greatly reduces the subjectivity of estimates. Finally, the AF analysis can be applied to any type of event series of sufficient size. For example, in Chapter 3, I was able to investigate Complexity Matching across multiple modalities.

5.4 Information and information transmission

For a functional hypothesis of information transmission to be empirically falsifiable, considerable work is needed to operationalize information and information transmission. Operationalizing these terms is important for empirical investigations and for theory building. In the case of simulations (Aquino et al., 2011; Beggs & Plenz, 2003), information transmission is the activation of another component in the network, e.g., neuron. How does this scale up to other levels of inquiry? For example, how can researchers operationalize information transmission during human interaction? Is there a 'ground truth' measure? These are important questions to be answered if a functional hypothesis of multiscale phenomena is to be regarded as tenable and empirically falsifiable by cognitive scientists.

Complexity Matching (West et al., 2008) was originally developed to formalize the relationship between complex networks and information transmission. Recent efforts in Psychology and Cognitive Science have made great strides in understanding the complex coordination patterns in human-human interactions (Abney, Paxton, et al., 2014; Coey et al., 2016; Fine et al., 2015; Marmelat & Delignières, 2012; Washburn, Kallen, Coey, Shockley, & Richardson, 2015) and human-environment interactions (Coey et al., 2014; Stephen et al., 2008, 2008; Marmelat & Delignières, 2012; Torre et al., 2013). Despite the progress in understanding the complex coordination patterns in human models, less work has focused on the main expectation of Complexity Matching: the consequence of information transmission. The results from Chapter 3 suggest that indirect information transmission, as measured by task performance, increases with greater matching of complex nested movement patterns across group members.

Future work in the Cognitive Sciences should focus on the operationalization of information and information transmission. For example, computational linguistics utilizes information theoretic measures of information in discourse analyses. One stronger test of the Complexity Matching hypothesis could test whether the convergence of complex human behaviors (e.g., hierarchical patterns of speech and/or movement) leads to increased rates of information in the linguistic structure of the conversation.

5.5 Development of hierarchical communicative structure

In Chapter 4, the production and convergence of hierarchical patterns in vocalizations were observed for infants and their caregivers across the first two years of life. Even at the youngest age of recording, 11-days-old, there was evidence for multiscale clustering of vocalizations. This observation points to the developmental origins of hierarchical nested communicative behavior. Moreover, I also observed that complexity matching between infant and caregiver vocal interactions was not fully explained by other coordination patterns like coincidence-based and rate-based vocal coordination.

There is a long history of studying coordination patterns in infant-caregiver interactions to gain insights into linguistic, social, emotional, and physiological aspects of development (Bateson, 1975; K. Bloom et al., 1987; Goldstein et al., 2003; Goldstein & Schwade, 2008; Jaffe et al., 2001; Kokkinaki & Kugiumutzakis, 2000; Nathani & Stark, 1996; Northrup & Iverson, 2015; Papoušek & Papoušek, 1989; Ramírez-Esparza et al., 2014; Warlaumont et al., 2014; Weisleder & Fernald, 2013). However, most of this research only focuses on one type of coordination pattern and level description. Such focus has undoubtedly progressed our understanding of the development of communication. Despite the benefits of focusing on a single level of description, multi-level theoretical frameworks can provide insights not afforded at a single level of description. The results from Chapter 4 provide an example of how taking a pluralistic approach (Abney, Dale, et al., 2014) to human interaction can yield insights into multiple areas of the phenomenon researchers are studying.

In addition to better understanding the role of multiple coordination patterns in infant-caregiver interaction, other patterns of unidirectional and bidirectional convergence were observed. Across the first two years of life, adults adapt the complexity of their vocalizations to their infants. It is possible that this convergence pattern is related to the fine-tuning hypothesis (Snow, 1977) that suggests that adults adapt their linguistic complexity during infant-directed speech to that of their child. Previous work on the fine-tuning hypothesis has observed that adults adapt their linguistic complexity to their infant across the age of the infant (Kunert et al., 2011; Snow, 1995; Sokolov, 1993). Showing a similar convergence pattern at a different communicative level provides support for the fine-tuning hypothesis but also adds a new dimension to this literature. Future research should focus on additional levels of communicative structure for converging evidence of the fine-tuning hypothesis.

5.6 Conclusion

This dissertation focused on testing the Complexity Matching hypothesis for human communication. I have presented three studies that focused on the production and convergence of hierarchical communicative behaviors across a variety of interactional contexts and populations. This dissertation also touched on new properties of communicative behavior such as the hierarchical and nested structure of speech and movement patterns. Future work should focus on the sensitivity of production and convergence of hierarchical behavioral patterns across more diverse contexts and populations along with seeking a better understanding of the necessary and sufficient conditions of optimal information transmission in human communication.

References

Abney, D. H., Dale, R., Yoshimi, J., Kello, C., Tylén, K., & Fusaroli, R. (2014). Joint perceptual decision-making: a case study in explanatory pluralism. *Frontiers in psychology*, *5*, 330–330.

Abney, D. H., Kello, C. T., & Warlaumont, A. S. (2015). Production and convergence of multiscale clustering in speech. *Ecological Psychology*, *27*(3), 222–235.

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

Abney, D. H., Paxton, A., Dale, R., & Kello, C. T. (2015). Movement dynamics reflect a functional role for weak coupling and role structure in dyadic problem solving. *Cognitive processing*, *16*(4), 325–332.

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*.

Allan, D. W. (1966). Statistics of atomic frequency standards. *Proceedings of the IEEE*, *54*(2), 221–230.

Altmann, E. G., Cristadoro, G., & Degli Esposti, M. (2012). On the origin of long-range correlations in texts. *Proceedings of the National Academy of Sciences*, *109*(29), 11582–11587.

Ambrose, S. E., VanDam, M., & Moeller, M. P. (2014). Linguistic input, electronic media, and communication outcomes of toddlers with hearing loss. *Ear and hearing*, *35*(2), 139.

Aquino, G., Bologna, M., Grigolini, P., & West, B. J. (2010). Beyond the death of linear response: 1/f optimal information transport. *Physical review letters*, *105*(4), 040601.

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.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of memory and language*, *59*(4), 390–412.

Bahrami, B., Olsen, K., Latham, P. E., Roepstorff, A., Rees, G., & Frith, C. D. (2010). Optimally interacting minds. *Science*, *329*(5995),

1081–1085.
Bak, P. (2013). *How nature works: the science of self-organized criticality.* Springer Science & Business Media.
Bateson, M. C. (1975). Mother-infant exchanges: The epigenesis of conversational interaction*. *Annals of the New York Academy of Sciences*, *263*(1), 101–113.
Beckner, C., Blythe, R., Bybee, J., Christiansen, M. H., Croft, W., Ellis, N. C., ... Schoenemann, T. (2009). Language is a complex adaptive system: Position paper. *Language learning*, *59*(s1), 1–26.
Beggs, J. M., & Plenz, D. (2003). Neuronal avalanches in neocortical circuits. *The Journal of neuroscience*, *23*(35), 11167–11177.
Bernieri, F. J., Reznick, J. S., & Rosenthal, R. (1988). Synchrony, pseudosynchrony, and dissynchrony: Measuring the entrainment process in mother-infant interactions. *Journal of personality and social psychology*, *54*(2), 243.
Bernstein, N. A. (1967). The co-ordination and regulation of movements.
Bloom, K., Russell, A., & Wassenberg, K. (1987). Turn taking affects the quality of infant vocalizations. *Journal of child language*, *14*(2), 211–227.
Bloom, L., Margulis, C., Tinker, E., & Fujita, N. (1996). Early conversations and word learning: Contributions from child and adult. *Child Development*, *67*(6), 3154–3175.
Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive psychology*, *18*(3), 355–387.
Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(6), 1482.
Brennan, S. E., & Hanna, J. E. (2009). Partner-specific adaptation in dialog. *Topics in Cognitive Science*, *1*(2), 274–291.
Cappella, J. N., & Planalp, S. (1981). Talk and silence sequences in informal conversations iii: Interspeaker influence. *Human Communication Research*, *7*(2), 117–132.
Caskey, M., Stephens, B., Tucker, R., & Vohr, B. (2011). Importance of parent talk on the development of preterm infant vocalizations. *Pediatrics*, *128*(5), 910–916.
Caskey, M., & Vohr, B. (2013). Assessing language and language environment of high-risk infants and children: a new approach. *Acta Paediatrica*, *102*(5), 451–461.
Clark, H. H., & Marshall, C. R. (1981). Definite reference and mutual knowledge.
Clauset, A., Shalizi, C. R., & Newman, M. E. (2009). Power-law distributions in empirical data. *SIAM review*, *51*(4), 661–703.
Coco, M. I., & Dale, R. (2013). Cross-recurrence quantification analysis of

categorical and continuous time series: an r package. *arXiv preprint arXiv:1310.0201*.

Coey, C. A., Washburn, A., Hassebrock, J., & Richardson, M. J. (2016). Complexity matching effects in bimanual and interpersonal syncopated finger tapping. *Neuroscience letters*, *616*, 204–210.

Coey, C. A., Washburn, A., & Richardson, M. J. (2014). Recurrence quantification as an analysis of temporal coordination with complex signals. In *Translational recurrences* (pp. 173–186). Springer.

Coupland, N. (1980). Style-shifting in a cardiff work-setting. *Language in Society*, *9*(01), 1–12.

Cox, R. F., & van Dijk, M. (2013). Microdevelopment in parent-child conversations: from global changes to flexibility. *Ecological Psychology*, *25*(3), 304–315.

Cummins, F., & Port, R. (1998). Rhythmic constraints on stress timing in english. *Journal of Phonetics*, *26*(2), 145–171.

Cutting, J. E., & Rosner, B. S. (1974). Categories and boundaries in speech and music*. *Perception & Psychophysics*, *16*(3), 564–570.

Dale, R., Fusaroli, R., Duran, N. D., & Richardson, D. C. (2014). The self-organization of human interaction.

Dale, R., Warlaumont, A. S., & Richardson, D. C. (2011). Nominal cross recurrence as a generalized lag sequential analysis for behavioral streams. *International Journal of Bifurcation and Chaos*, *21*(04), 1153–1161.

Dikker, S., Silbert, L. J., Hasson, U., & Zevin, J. D. (2014). On the same wavelength: Predictable language enhances speaker–listener brain-to-brain synchrony in posterior superior temporal gyrus. *The Journal of Neuroscience*, *34*(18), 6267–6272.

Feldman, R. (2007). Parent–infant synchrony and the construction of shared timing; physiological precursors, developmental outcomes, and risk conditions. *Journal of Child Psychology and Psychiatry*, *48*(3-4), 329–354.

Ferrer i Cancho, R., Solé, R. V., & Köhler, R. (2004). Patterns in syntactic dependency networks. *Physical Review E*, *69*(5), 051915.

Fine, J. M., Likens, A. D., Amazeen, E. L., & Amazeen, P. G. (2015). Emergent complexity matching in interpersonal coordination: Local dynamics and global variability.

Franklin, B., Warlaumont, A. S., Messinger, D., Bene, E., Nathani Iyer, S., Lee, C.-C., ... Oller, D. K. (2014). Effects of parental interaction on infant vocalization rate, variability and vocal type. *Language Learning and Development*, *10*(3), 279–296.

Fuchs, A., Jirsa, V. K., Haken, H., & Kelso, J. S. (1996). Extending the hkb model of coordinated movement to oscillators with different eigenfrequencies. *Biological cybernetics*, *74*(1), 21–30.

Fusaroli, R., Abney, D., Bahrami, B., Kello, C., & Tylén, K. (2013). Conversation, coupling and complexity: matching scaling laws predict performance in a joint decision task. In *Poster presented at the 35th annual conference of the cognitive science society.*

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*, 0956797612436816.

Fusaroli, R., Konvalinka, I., & Wallot, S. (2014). Analyzing social interactions: the promises and challenges of using cross recurrence quantification analysis. In *Translational recurrences* (pp. 137–155). Springer.

Fusaroli, R., & Tylén, K. (2015). Investigating conversational dynamics: Interactive alignment, interpersonal synergy, and collective task performance. *Cognitive science*.

Garrod, S., & Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition*, *27*(2), 181–218.

Gilkerson, J., & Richards, J. A. (2008). The lena natural language study. *Boulder, CO: LENA Foundation. Retrieved March*, *3*, 2009.

Gilkerson, J., Richards, J. A., & Topping, K. J. (2015). The impact of book reading in the early years on parent–child language interaction. *Journal of Early Childhood Literacy*, 1468798415608907.

Giveans, M. R., Pelzer, C., Smith, A., Shockley, K., & Stoffregen, T. A. (2008). Postural support for personal performance and interpersonal coordination. In *Journal of sport & exercise psychology* (Vol. 30, pp. S20–S20).

Goldstein, M. H., King, A. P., & West, M. J. (2003). Social interaction shapes babbling: Testing parallels between birdsong and speech. *Proceedings of the National Academy of Sciences*, *100*(13), 8030–8035.

Goldstein, M. H., & Schwade, J. A. (2008). Social feedback to infants' babbling facilitates rapid phonological learning. *Psychological Science*, *19*(5), 515–523.

Goldstein, M. H., Schwade, J. A., & Bornstein, M. H. (2009). The value of vocalizing: Five-month-old infants associate their own noncry vocalizations with responses from caregivers. *Child development*, *80*(3), 636–644.

Goldstein, M. H., & West, M. J. (1999). Consistent responses of human mothers to prelinguistic infants: the effect of prelinguistic repertoire size. *Journal of Comparative Psychology*, *113*(1), 52.

Greenwood, C. R., Thiemann-Bourque, K., Walker, D., Buzhardt, J., & Gilkerson, J. (2011). Assessing children's home language environments using automatic speech recognition technology. *Communication Disorders Quarterly*, *32*(2), 83–92.

Haken, H., Kelso, J. S., & Bunz, H. (1985). A theoretical model of phase transitions in human hand movements. *Biological cybernetics*, *51*(5), 347–356.

Harder, S., Lange, T., Hansen, G. F., Væver, M., & Køppe, S. (2015). A longitudinal study of coordination in mother–infant vocal interaction from age 4 to 10 months. *Developmental psychology*, *51*(12), 1778.

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young american children.* Paul H Brookes Publishing.

Hasson, U., Ghazanfar, A. A., Galantucci, B., Garrod, S., & Keysers, C. (2012). Brain-to-brain coupling: a mechanism for creating and sharing a social world. *Trends in cognitive sciences*, *16*(2), 114–121.

Healey, P. G. (2008). Interactive misalignment: The role of repair in the development of group sub-languages. *Language in Flux. College Publications*, *212*.

Healey, P. G., Purver, M., & Howes, C. (2014). Divergence in dialogue. *PLoS ONE*, *9*(6).

Hilbrink, E. E., Gattis, M., & Levinson, S. C. (2015). Early developmental changes in the timing of turn-taking: a longitudinal study of mother–infant interaction. *Frontiers in psychology*, *6*.

Holden, J. G., Van Orden, G. C., & Turvey, M. T. (2009). Dispersion of response times reveals cognitive dynamics. *Psychological review*, *116*(2), 318.

Howes, C., Healey, P. G., & Purver, M. (2010). Tracking lexical and syntactic alignment in conversation. In *Proceedings of the 20th annual conference of the cognitive science society* (pp. 2004–2009).

Hsu, H.-C., Fogel, A., & Messinger, D. S. (2001). Infant non-distress vocalization during mother-infant face-to-face interaction: Factors associated with quantitative and qualitative differences. *Infant behavior and development*, *24*(1), 107–128.

Ireland, M. E., Slatcher, R. B., Eastwick, P. W., Scissors, L. E., Finkel, E. J., & Pennebaker, J. W. (2011). Language style matching predicts relationship initiation and stability. *Psychological Science*, *22*(1), 39–44.

Jaeger, T. F., & Snider, N. (2008). Implicit learning and syntactic persistence: Surprisal and cumulativity. In *Proceedings of the cognitive science society conference* (pp. 1061–1066).

Jaffe, J., Beebe, B., Feldstein, S., Crown, C. L., Jasnow, M. D., Rochat, P., & Stern, D. N. (2001). Rhythms of dialogue in infancy: Coordinated timing in development. *Monographs of the society for research in child development*, i–149.

Johnson, K., Caskey, M., Rand, K., Tucker, R., & Vohr, B. (2014). Gender differences in adult-infant communication in the first months of life. *Pediatrics*, *134*(6), e1603–e1610.

Johnson, S. C. (1967). Hierarchical clustering schemes. *Psychometrika*, *32*(3), 241–254.

Keller, P. E., & Repp, B. H. (2004). When two limbs are weaker than one: Sensorimotor syncopation with alternating hands. *The Quarterly Journal of Experimental Psychology Section A*, *57*(6), 1085–1101.

Kello, C. T., & Beltz, B. C. (2009). Scale-free networks in phonological and orthographic wordform lexicons. *Approaches to phonological complexity*, 171–190.

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.

Kelso, J. (1981). On the oscillatory basis of movement. In *Bulletin of the psychonomic society* (Vol. 18, pp. 63–63).

Kelso, J. (1997). *Dynamic patterns: The self-organization of brain and behavior*. MIT press.

Kokkinaki, T., & Kugiumutzakis, G. (2000). Basic aspects of vocal imitation in infant-parent interaction during the first 6 months. *Journal of reproductive and infant psychology*, *18*(3), 173–187.

Kugler, P. N. (1987). *Information, natural law, and the self-assembly of rhythmic movement*. Routledge.

Kunert, R., Fernández, R., & Zuidema, W. (2011). Adaptation in child directed speech: Evidence from corpora. *Proc. SemDial*, 112–119.

Levelt, W. J. (1969). The perception of syntactic structure. *Heymans Bulletins*.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of experimental psychology*, *54*(5), 358.

Louwerse, M. M., Dale, R., Bard, E. G., & Jeuniaux, P. (2012). Behavior matching in multimodal communication is synchronized. *Cognitive Science*, *36*(8), 1404–1426.

Luque, J., Luque, B., & Lacasa, L. (2015). Scaling and universality in the human voice. *Journal of The Royal Society Interface*, *12*(105), 20141344.

Lynch, M. P., Oller, D. K., Steffens, M. L., & Buder, E. H. (1995). Phrasing in prelinguistic vocalizations. *Developmental psychobiology*, *28*(1), 3–25.

Mandelbrot, B. B. (1983). *The fractal geometry of nature* (Vol. 173). Macmillan.

Manson, J. H., Bryant, G. A., Gervais, M. M., & Kline, M. A. (2013). Convergence of speech rate in conversation predicts cooperation. *Evolution and Human Behavior*, *34*(6), 419–426.

Marmelat, V., & Delignières, D. (2012). Strong anticipation: complexity matching in interpersonal coordination. *Experimental Brain Re-*

search, *222*(1-2), 137–148.

Marwan, N., Romano, M. C., Thiel, M., & Kurths, J. (2007). Recurrence plots for the analysis of complex systems. *Physics reports*, *438*(5), 237–329.

McClave, E. Z. (2000). Linguistic functions of head movements in the context of speech. *Journal of pragmatics*, *32*(7), 855–878.

Miller, G. A., & Chomsky, N. (1963). Finitary models of language users.

Mills, G. J. (2014). Dialogue in joint activity: complementarity, convergence and conventionalization. *New ideas in psychology*, *32*, 158–173.

Mitchell, M. (2009). *Complexity: A guided tour*. Oxford University Press.

Mitzenmacher, M. (2004). A brief history of generative models for power law and lognormal distributions. *Internet mathematics*, *1*(2), 226–251.

Nathani, S., & Stark, R. E. (1996). Can conditioning procedures yield representative infant vocalizations in the laboratory? *First language*, *16*(48), 365–387.

Neumann, R., & Strack, F. (2000). " mood contagion": the automatic transfer of mood between persons. *Journal of personality and social psychology*, *79*(2), 211.

Niederhoffer, K. G., & Pennebaker, J. W. (2002). Linguistic style matching in social interaction. *Journal of Language and Social Psychology*, *21*(4), 337–360.

Northrup, J. B., & Iverson, J. M. (2015). Vocal coordination during early parent–infant interactions predicts language outcome in infant siblings of children with autism spectrum disorder. *Infancy*, *20*(5), 523–547.

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

Oller, D. K., Eilers, R. E., Basinger, D., Steffens, M. L., & Urbano, R. (1995). Extreme poverty and the development of precursors to the speech capacity. *First Language*, *15*(44), 167–187.

Oller, D. K., Niyogi, P., Gray, S., Richards, J., Gilkerson, J., Xu, D., ... Warren, S. (2010). Automated vocal analysis of naturalistic recordings from children with autism, language delay, and typical development. *Proceedings of the National Academy of Sciences*, *107*(30), 13354–13359.

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.

Papoušek, M., & Papoušek, H. (1989). Forms and functions of vocal matching in interactions between mothers and their precanonical infants. *First language*, *9*(6), 137–157.

Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America*, *119*(4), 2382–2393.

Pardo, J. S. (2013). Measuring phonetic convergence in speech production. *Frontiers in Psychology*, *4*.

Paxton, A., & Dale, R. (2013a). Argument disrupts interpersonal synchrony. *The Quarterly Journal of Experimental Psychology*, *66*(11), 2092–2102.

Paxton, A., & Dale, R. (2013b). Frame-differencing methods for measuring bodily synchrony in conversation. *Behavior Research Methods*, *45*(2), 329–343.

Pickering, M. J., & Branigan, H. P. (1998). The representation of verbs: Evidence from syntactic priming in language production. *Journal of Memory and Language*, *39*(4), 633–651.

Pickering, M. J., & Branigan, H. P. (1999). Syntactic priming in language production. *Trends in cognitive sciences*, *3*(4), 136–141.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and brain sciences*, *27*(02), 169–190.

Ramenzoni, V. C., Riley, M. A., Shockley, K., & Baker, A. A. (2012). Interpersonal and intrapersonal coordinative modes for joint and single task performance. *Human Movement Science*, *31*(5), 1253–1267.

Ramírez-Esparza, N., García-Sierra, A., & Kuhl, P. K. (2014). Look who's talking: speech style and social context in language input to infants are linked to concurrent and future speech development. *Developmental science*, *17*(6), 880–891.

Rescorla, L., & Ratner, N. B. (1996). Phonetic profiles of toddlers with specific expressive language impairment (sli-e). *Journal of Speech, Language, and Hearing Research*, *39*(1), 153–165.

Richardson, D. C., & Dale, R. (2005). Looking to understand: The coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension. *Cognitive science*, *29*(6), 1045–1060.

Richardson, M. J., Harrison, S. J., Kallen, R. W., Walton, A., Eiler, B. A., Saltzman, E., & Schmidt, R. (2015). Self-organized complementary joint action: Behavioral dynamics of an interpersonal collision-avoidance task. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(3), 665.

Richardson, M. J., Marsh, K. L., Isenhower, R. W., Goodman, J. R., & Schmidt, R. C. (2007). Rocking together: Dynamics of intentional and unintentional interpersonal coordination. *Human movement science*, *26*(6), 867–891.

Richardson, M. J., Marsh, K. L., & Schmidt, R. (2005). Effects of visual and verbal interaction on unintentional interpersonal coordination. *Journal of Experimental Psychology: Human Perception and Perfor-*

mance, *31*(1), 62.

Riley, M. A., Richardson, M. J., Shockley, K., & Ramenzoni, V. C. (2011). Interpersonal synergies. *Frontiers in psychology*, *2*, 38.

Riley, M. A., & Turvey, M. T. (2002). Variability and determinism in motor behavior. *Journal of motor behavior*, *34*(2), 99–125.

Sales-Pardo, M., Guimera, R., Moreira, A. A., & Amaral, L. A. N. (2007). Extracting the hierarchical organization of complex systems. *Proceedings of the National Academy of Sciences*, *104*(39), 15224–15229.

Schmidt, R. C., Carello, C., & Turvey, M. T. (1990). Phase transitions and critical fluctuations in the visual coordination of rhythmic movements between people. *Journal of experimental psychology: human perception and performance*, *16*(2), 227.

Schmidt, R. C., Morr, S., Fitzpatrick, P., & Richardson, M. J. (2012). Measuring the dynamics of interactional synchrony. *Journal of Nonverbal Behavior*, *36*(4), 263–279.

Schmidt, R. C., & Richardson, M. J. (2008). Dynamics of interpersonal coordination. In *Coordination: Neural, behavioral and social dynamics* (pp. 281–308). Springer.

Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: bodies and minds moving together. *Trends in cognitive sciences*, *10*(2), 70–76.

Shockley, K., Baker, A. A., Richardson, M. J., & Fowler, C. A. (2007). Articulatory constraints on interpersonal postural coordination. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(1), 201.

Shockley, K., Richardson, D. C., & Dale, R. (2009). Conversation and coordinative structures. *Topics in Cognitive Science*, *1*(2), 305–319.

Shockley, K., Santana, M.-V., & Fowler, C. A. (2003). Mutual interpersonal postural constraints are involved in cooperative conversation. *Journal of Experimental Psychology: Human Perception and Performance*, *29*(2), 326.

Simon, H. A. (1977). The organization of complex systems. In *Models of discovery* (pp. 245–261). Springer.

Snow, C. E. (1977). Mothers' speech research: From input to interaction. *Talking to children: Language input and acquisition*, 31–49.

Snow, C. E. (1989). Understanding social interaction and language acquisition; sentences are not enough.

Snow, C. E. (1995). Issues in the study of input: Finetuning, universality, individual and developmental differences, and necessary causes. *The handbook of child language*, 180–193.

Soderstrom, M., & Wittebolle, K. (2013). When do caregivers talk? the influences of activity and time of day on caregiver speech and child vocalizations in two childcare environments. *PloS one*, *8*(11), e80646.

Sokolov, J. L. (1993). A local contingency analysis of the fine-tuning

hypothesis. *Developmental psychology*, *29*(6), 1008.
Spivey, M. (2007). The continuity of mind.
Stanley, H. E. (1987). Introduction to phase transitions and critical phenomena. *Introduction to Phase Transitions and Critical Phenomena, by H Eugene Stanley, pp. 336. Foreword by H Eugene Stanley. Oxford University Press, Jul 1987. ISBN-10: 0195053168. ISBN-13: 9780195053166*, *1*.
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.
Stephen, D. G., & Dixon, J. A. (2011). Strong anticipation: Multifractal cascade dynamics modulate scaling in synchronization behaviors. *Chaos, Solitons & Fractals*, *44*(1), 160–168.
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.
Stephen, D. G., Stepp, N., Dixon, J. A., & Turvey, M. (2008). Strong anticipation: Sensitivity to long-range correlations in synchronization behavior. *Physica A: Statistical Mechanics and its Applications*, *387*(21), 5271–5278.
Stephens, G. J., Silbert, L. J., & Hasson, U. (2010). Speaker–listener neural coupling underlies successful communication. *Proceedings of the National Academy of Sciences*, *107*(32), 14425–14430.
Stergiou, N., & Decker, L. M. (2011). Human movement variability, nonlinear dynamics, and pathology: is there a connection? *Human movement science*, *30*(5), 869–888.
Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., ... others (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences*, *106*(26), 10587–10592.
Strogatz, S. H., & Mirollo, R. E. (1991). Stability of incoherence in a population of coupled oscillators. *Journal of Statistical Physics*, *63*(3-4), 613–635.
Suchman, L. (2007). *Human-machine reconfigurations: Plans and situated actions*. Cambridge University Press.
Suci, G., Ammon, P., & Gamlin, P. (1967). The validity of the probe-latency technique for assessing structure in language. *Language and speech*, *10*(2), 69.
Suskind, D. L., Leffel, K. R., Graf, E., Hernandez, M. W., Gunderson, E. A., Sapolich, S. G., ... Levine, S. C. (2015). A parent-directed language intervention for children of low socioeconomic status: a randomized controlled pilot study. *Journal of child language*, 1–41.

Swenson, R., & Turvey, M. T. (1991). Thermodynamic reasons for perception–action cycles. *Ecological Psychology*, *3*(4), 317–348.

Thelen, E., & Smith, L. (1994). *A dynamic systems approach to the development of perception and action.* MIT Press Cambridge, MA.

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*(04), 565–595.

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.

Turalska, M., West, B. J., & Grigolini, P. (2011). Temporal complexity of the order parameter at the phase transition. *Physical Review E*, *83*(6), 061142.

Turvey, M. T. (1990). Coordination. *American psychologist*, *45*(8), 938.

Turvey, M. T. (2007). Action and perception at the level of synergies. *Human movement science*, *26*(4), 657–697.

VanDam, M. (2014). Acoustic characteristics of the clothes used for a wearable recording device. *The Journal of the Acoustical Society of America*, *136*(4), EL263–EL267.

VanDam, M., Oller, D. K., Ambrose, S. E., Gray, S., Richards, J. A., Xu, D., ... Moeller, M. P. (2015). Automated vocal analysis of children with hearing loss and their typical and atypical peers. *Ear and hearing*, *36*(4), e146–e152.

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

Voss, R. F., & Clarke, J. (1978). "1/f noise"in music: Music from 1/f noise. *The Journal of the Acoustical Society of America*, *63*(1), 258–263.

Warlaumont, A. S., Oller, D. K., Dale, R., Richards, J. A., Gilkerson, J., & Xu, D. (2010). Vocal interaction dynamics of children with and without autism. In *Proceedings of the 32nd annual conference of the cognitive science society* (pp. 121–126).

Warlaumont, A. S., Richards, J. A., Gilkerson, J., & Oller, D. K. (2014). A social feedback loop for speech development and its reduction in autism. *Psychological science*, 0956797614531023.

Warren, S. F., Gilkerson, J., Richards, J. A., Oller, D. K., Xu, D., Yapanel, U., & Gray, S. (2010). What automated vocal analysis reveals about the vocal production and language learning environment of young children with autism. *Journal of autism and developmental disorders*, *40*(5), 555–569.

Washburn, A., Kallen, R. W., Coey, C. A., Shockley, K., & Richardson, M. J. (2015). Harmony from chaos? perceptual-motor delays enhance behavioral anticipation in social interaction. *Journal of experimental* *psychology: human perception and performance*, *41*(4), 1166.

Webb, J. T. (1969). Subject speech rates as a function of interviewer behaviour. *Language and Speech*, *12*(1), 54–67.

Weisleder, A., & Fernald, A. (2013). Talking to children matters early language experience strengthens processing and builds vocabulary. *Psychological science*, *24*(11), 2143–2152.

West, B. J., & Deering, W. (1994). Fractal physiology for physicists: Lévy statistics. *Physics Reports*, *246*(1-2), 1–100.

West, B. J., Geneston, E. L., & Grigolini, P. (2008). Maximizing information exchange between complex networks. *Physics Reports*, *468*(1), 1–99.

Wijnants, M. L., Bosman, A. M., Hasselman, F., Cox, R. F., & Van Orden, G. C. (2009). 1/f scaling in movement time changes with practice in precision aiming. *Nonlinear dynamics, psychology, and life sciences*, *13*(1), 79.

Xu, D., Yapanel, U., & Gray, S. (2009). *Reliability of the lenatm language environment analysis system in young children's natural home environment* (Tech. Rep.). LENA Foundation Technical Report LTR-05-02). Retrieved from http://www. lenafoundation. org/TechReport. asp x/Reliability/LTR-05-2.

Xu, D., Yapanel, U., Gray, S., & Baer, C. (2008). The lena™ language environment analysis system: The interpretive time segments (its) file. *LENA Research Foundation Technical Report LTR-04-2*.

Zipf, G. K. (1949). Human behavior and the principle of least effort.