Network Analysis of Multimodal, Multiscale Coordination in Dyadic Problem Solving

Alexandra Paxton (paxton.alexandra@gmail.com)
Drew H. Abney (dabney@ucmerced.edu)
Christopher T. Kello (ckello@ucmerced.edu)
Rick Dale (rdale@ucmerced.edu)
Cognitive & Information Sciences, University of California, Merced
Merced, CA 95343 USA

Abstract

A recent trend in dyadic interaction research utilizes multiple modalities to better understand phenomena encompassing *behavior matching* (e.g., synchrony, alignment). Concurrent research has focused on a complementary framework of interaction, assessing the matching of power law distributions of behavior across two people: *complexity matching*. While both frameworks provide useful insights into dyadic interaction, they have done so independent of one another. We visualize the multimodal, multiscale coordination of dyads engaged in a tower-building task as networks based on the analyses of behavioral and complexity matching in speech and movement. We find that network strength relates to task performance and that high-performing dyads have weaker network strength, which we argue opens up more degrees of freedom affording more flexibility in the dyadic system.

Keywords: communication; complexity matching; convergence; dyadic interaction; interpersonal synchrony; networks

Introduction

Interpersonal communication lives across a number of timescales. During face-to-face interaction, our signals to one another can last from milliseconds to hours, and smooth interaction requires an effective juggling of incoming and outgoing signals. We readily perceive and quickly respond to more obvious signals like facial expressions and linguistic information, but we also influence one another in more subtle ways. From language (Brennan & Clark, 1996; Garrod & Pickering, 2004) and emotion (Hatfield, Cacioppo, & Rapson, 1995) to neural patterns (Stevens, Silbert, & Hasson, 2010) and physiological signals (Helm, Sbarra, & Ferrer, 2012), research on *interpersonal convergence* (or *coordination*, *entrainment*, or *synchrony*) highlights ways in which we influence and are influenced by those with whom we interact. This often builds on a large body of joint action literature (Clark, 1996), exploring how we come to work together.

Previous research on convergence has tended to focus on specific behaviors or patterns (e.g., Louwerse, Dale, Bard, & Jeuniaux, 2012), but an emerging and exciting focus instead investigates interpersonal convergence of the statistical patterns of behavior. This focus is generally called *complexity matching* (e.g., Abney, Paxton, Kello, & Dale, under revision; Marmelat & Deligniéres, 2012), contrasting with the *behavior matching* prevalent in traditional convergence research. While many behaviors targeted for study in behavior matching are overt and perceptible to others during interactions, complexity matching focuses on convergence at the *distributional* level of conversational properties.

Complexity matching is a special case of convergence of distribution-level patterns of behaviors: It captures the matching of behaviors that follow power law distributions. Power law distributions are indicative of multiscale variations and are exhibited by complex systems (Sales-Pardo, Guimera, Moreira, & Amaral, 2007), hence the use of the term *complexity matching* to quantify the matching of these properties across two people in an interaction. The notion of complexity matching of two humans interacting is suggested by recent research showing that when the power law distributions of interacting complex systems match, optimal information transmission occurs (West, Geneston, & Grigolini, 2008). Therefore, we hypothesize that when the behaviors of two humans follow a power law distribution, the degree of matching between these quantitative patterns might reflect properties of the interaction like information flow, context, and valence.

While the framework of behavior matching quantifies the one-to-one matching of behaviors during an interaction (e.g., gaze patterns; Louwerse et al., 2012), complexity matching quantifies the degree to which particular statistical patterns (e.g., patterns of behavior that are power-law distributed) match throughout an entire interaction. Thus, behavior matching and complexity matching are complementary measures of interpersonal convergence. In the present study, we use both behavior and complexity matching to create networks of speech and movement in dyadic interaction during a cooperative task.

The Present Study

Just as recent trends in interpersonal interaction research combine multiple communication channels into multimodal analyses (e.g., Louwerse et al., 2012; Paxton & Dale, 2013b), our understanding of interaction can significantly benefit from integrating analyses on multiple time scales. One promising way to do this may be through combining behavior matching with complexity matching. Such analyses combine these two methods to permit investigation of interlocutors' tendencies to (1) exhibit similar behaviors over time (behavioral matching), (2) organize behaviors similarly in time (complexity matching), or (3) a combination of these two. Two people may not only match moment-to-moment behaviors (i.e., behavior matching); they may also exhibit behaviors in characteristic ways over a longer time course (i.e., complexity matching).

The present study is aimed at testing this possibility while contributing to work on interpersonal convergence and task performance. Here, we analyze dyadic interaction during an engaging but somewhat challenging collaborative task: constructing towers out of marshmallows and spaghetti (e.g., Wujec, 2010). Previous work suggests that behavior matching may improve collaborative performance (e.g., Fusaroli et al., 2012; Valdesolo, Ouyang, & DeSteno, 2010), and we therefore anticipate that performance will be positively related to increased behavior and complexity matching. We also believe that behavior matching and complexity matching will be closely related to one another, as suggested by some parallel findings across separate studies of behavior matching (Paxton & Dale, 2013b) and complexity matching (Abney et al., under revision).

We employ network-style visualizations (Paxton & Dale, 2013b) to showcase the interconnectivity of these data as a comprehensive framework for integrating multiple modalities and scales of convergence of the dyadic-level system in a relatively intuitive graph. As we describe in more detail below, we compare networks of dyadic-level variables by partitioning data according to task performance. The network visualization focuses on the difference in network strengths of high- and low-performing dyads, which facilitates investigations of the interaction network. Using this method, we hypothesize that high-performing dyads should have lower network strengths compared with low-performing dyads, as effective cooperative performance in complex tasks may require flexible shifting over a range of interaction patterns to meet changing task demands.

Method

Participants

Twenty-four undergraduate students (mean age=19.7 years) at the University of California, Merced participated as dyads in return for extra course credit. Participants signed up for time slots anonymously and were unable to see partners' identities beforehand. Dyads included female-female (n=5), male-male (n=3), and mixed sex pairings (n=4).

Materials and Procedure

Following a brief demographics survey, participants were asked to sit in one of two chairs near a table. Seating arrangement was not programmatically controlled, and participants arranged themselves without experimenter direction. The two chairs and table were oriented such that the chairs were placed adjacent to each other, with the table rotated 45° in line of sight of the camcorder.

Once seated, the participants were given task instructions. Participants were told to construct the tallest tower structure possible within 15 minutes using only the materials provided: one box (~10 oz) of marshmallows and one box (~1 lb) of raw spaghetti. Importantly, only one participant seated on the right was allowed to touch the marshmallows, and only the participant seated on the left was allowed to touch the spaghetti. They were not allowed to use partial pieces of materials, and any materials that broke during construction were to be immediately removed from the tower. Participants were permitted to talk freely during construction. After answering any questions, the experimenters started the task.

Experimenters provided 5-minute and 1-minute warnings. After the time limit expired, the experimenters recorded the height and weight of the tower. Participants were separated and rated perceptions of the roles of marshmallow and spaghetti holders ("for most people who complete this task") on a *1 (mostly passive)* to *4 (mostly dominant)* scale. This enabled us to investigate how much participants believed power should be distributed during the task.

Apparatus and Data Preparation

Interactions were recorded on a tripod-mounted Canon Vixia HF M31 HD Camcorder. Audio for each participant was recorded separately (44kHz sample rate) with two Shure Beta 54 supercardiod microphone headsets, an M-Audio MobilePre recording interface, and Audacity software. Two audio files (1 per participant) were recorded per conversation. Video and audio files were synchronized with Apple iMovie software and truncated to contain only interactions occurring during the task.

The video files were then analyzed using a frame-differencing method (FDM) to obtain time series of standardized movement scores for each participant based on changes in pixels from frame to frame in a recorded interaction (see Paxton & Dale, 2013a), such that higher numbers in the time series indicated higher amounts of overall movement for that participant. The audio files for each participant were analyzed by using the Audacity "sound finder" to locate acoustic onset/offset intervals. The threshold of acoustic intensity was set at -30dB for all audio files. Due to low acoustic intensity values for a majority of recordings from the left channel, acoustic intensity was amplified by 6dB for all audio files from the left channel. Movement and speech data were chosen for analysis due to their high salience as communication channels during interpersonal interaction and the ability to collect both unobtrusively during interaction, facilitating naturalistic interaction while still collecting multimodal data.

Analyses and Results

To better understand the relations between different timescales of convergence, we chose to model our data in a fully interconnected network-style visualization (e.g., Paxton & Dale, 2013b). Each node in this network is a single time series or type of data, and each connection strength is the effect size of the linear model predicting one node to another. All data were standardized prior to being entered into the models, allowing estimates to be interpreted as effect sizes (Keith, 2005).

We present a network model of our data comparing patterns of multimodal, multiscale convergence exhibited in high- and low-performing dyads. This network models the interaction at the dyadic level, with all metrics calculated across the entire dyad. Before discussing this network, we first detail the nodes included.

Behavior Matching: Cross-Correlational Analyses

Behavior matching was assessed with cross-correlation of participants' data, which allowed us to explore patterns of influence between participants at various time lags. Cross-correlation shifted data at specified lags (e.g., comparing time t of one participant with time t+1 of the other) to calculate the extent of correlation between two time series within given windows. We calculated the cross-correlation coefficients between participants within dyads within +/- 3 seconds (at 8Hz) for each modality, resulting in a single series of cross-correlation coefficients per dyad for movement and for speech. The movement cross-correlation analyses used the standardized movement time series from the FDM analysis; the audio cross-correlation coefficients used the on/off speech state time series for each participant.

Consistent with previous findings (e.g., Louwerse et al., 2012), we found evidence to support time-locked speech and movement behavior matching between participants (ps $< .001$) across the interactions.[1] To retain the temporal qualities of the cross-correlation coefficients, we created interaction terms between these cross-correlation coefficients and time lag that serve as the nodes for the behavior matching (BM) in our networks. These new variables measured the degree of behavior matching occurring in a small window of time around simultaneous behavior while still weighting behavior matching that occurs *in time* most heavily.

Complexity Matching: Allan Factor Analyses

To complement the behavior matching analyses, the distributional information from the movement and speech behaviors was matched across participants in a dyad. Movement and speech behaviors have been observed to follow power law-like distributions (Abney et al., under revision) and that these distributions match across people in various types of interactions. Allan Factor (AF) analysis (Allan, 1966) was used to estimate the correlated clustering of behavioral events of each type across multiple time scales. The AF analysis estimated the variance of behavior events (e.g., onsets of movement or speech) at particular time scales and computed the correlation estimate (α) across those multiple time scales. A scaling relation of behavioral events was evidenced when α~1; α~0 was considered a Poisson process. This scaling relation is a power law and relates to the clustering of behavioral activity across multiple time scales (from 160ms to 10s).

The AF analysis is a point process analysis and requires binary spike trains of events and nonevents. For the movement data, binary spike trains were computed from the original z-score movement series derived from the FDM described earlier. Onset/offset states (coded as *1*) and operationalized as movement that rose or fell above or below the mean (respectively); all other states were coded as

[1] Based on separate linear mixed-effects models predicting cross-correlation coefficients of movement and speech with lag (+/- 3 sec) as fixed effect and with dyad and participant as non-nested random effects with fully specified random slopes. Movement: $ß = -.67$, $p < .001$. Speech: $ß = -.54$, $p < .001$. Dyads experience the highest cross-correlation values at synchrony for both.

0. Binary spike trains for speech were computed from the interval-level data (onset states = *1*; all other states = *0*).

To quantify complexity matching (or the matching of the estimates between participants in a dyad), we calculated an absolute difference of the AF functions for participants in each dyad. These absolute difference values were summed over the multiple timescales to create a single value that captures the degree of matching. The summed absolute difference value was considered the metric of complexity matching (CM). Unlike BM, *smaller* CM values were interpreted as *higher* rates of convergence.

Behavior Matching and Complexity Matching

The ability to test the relationship between two scales of convergence – behavior matching and complexity matching – provides a more comprehensive look into how dyads organize speech and movement behaviors across the problem-solving task than either alone. For example, participants' movements could phase in and out of synchrony but could nevertheless remain coordinated at the level of complexity matching (e.g., highly regular turn-taking structure). Past work has either studied behavior matching or complexity matching in completely separate studies, often in different domains. In the present work, we are able to leverage both techniques' strengths to better understand the multimodal, multiscale interaction structure.

Before creating the network, we tested the relationship between the two convergence patterns for each modality and across both halves of the interaction. Results suggested there were no reliable relationships between BM_{speech} and CM_{speech} for the first (β = -.004, p = .373) or the second half of the interactions (β = .003, p = .469). However, reliable relationships were found between BM_{mov} and CM_{mov} for the first (β = -.109, $p < .009$) and second halves of the interactions (β = -.133, $p < .001$). For participants' movement – but not their speech – behavior matching increased concurrently with an increase in complexity matching, a trend that increased during the second half of the interaction. (Again, greater convergence should be reflected in *positive* BM values and *negative* CM values.)

Performance and Social Data

We next analyzed performance and social measures. The performance metric – a ratio of height to weight of the tower – captured performance relative to materials used.[2] Notably, a linear regression confirmed significant relationships between performance (as a continuous variable) and measures of convergence. Improvements in performance were reliably predicted by BM_{speech} (β = .087, p = .04), BM_{mov} (β = .163, $p < .001$), and CM_{speech} (β = -.470, $p < .001$) but not CM_{mov} (β = .045, p = .274). We calculated a median split to obtain high- and low-performing groups.

We also created a dyadic-level variable operationalizing the perception of role distribution. For each dyad, we calculated a dyad-level dominance score for the spaghetti holder as a sum of each participant's dominance rating for the spaghetti holder, divided by the sum of the participants' individual perceptions of the marshmallow holder's dominance. This variable – which we call "role distribution" – tapped into dyads' expectations about role division: Higher values indicated the dyad overall endorsed a stronger leader-follower dynamic in the task, while lower values implied a more egalitarian expectation for the interaction.

Generating Network Visualizations

For the networks, we divided the data into two equal groups along the performance variable. We performed a series of linear models, each predicting one node by one other node until fully interconnected network was complete. The resulting effect sizes for each model were used as the connection strengths between nodes.

For a broad measure of network strength, we computed the average effect size for each network using the absolute values of effect sizes, allowing averages to be agnostic to positive versus negative effect sizes. We chose to use the absolute rather than the signed values due to the differences in BM and CM metrics: Higher convergence would yield a higher BM metric but a lower CM metric. All effect sizes obtained from the models were included in the calculation of the network strengths, regardless of p-value, to provide a full estimate of all connections: Connections not significant at $p < .05$ had an average absolute connection strength of .04 (range = |.002-.09|) and were equally distributed across the networks. Additional calculations of network strengths using only significant ($p < .05$) connections and only significant-to-marginal ($p < .1$) connections followed patterns similar to those using all connections.

It is important to note that, while these visualizations offer an inherently interesting look at the data, the significance levels are not essential to the inferences we make about the network structures. Effect sizes from the linear models then become data for the comparative network analysis. The connection strengths are summed to obtain a single measure of network strength, and all connections of each individual network are fed into t-tests, which constitute our comparison between the networks.

[2] This ratio was used due to low variance of height alone.

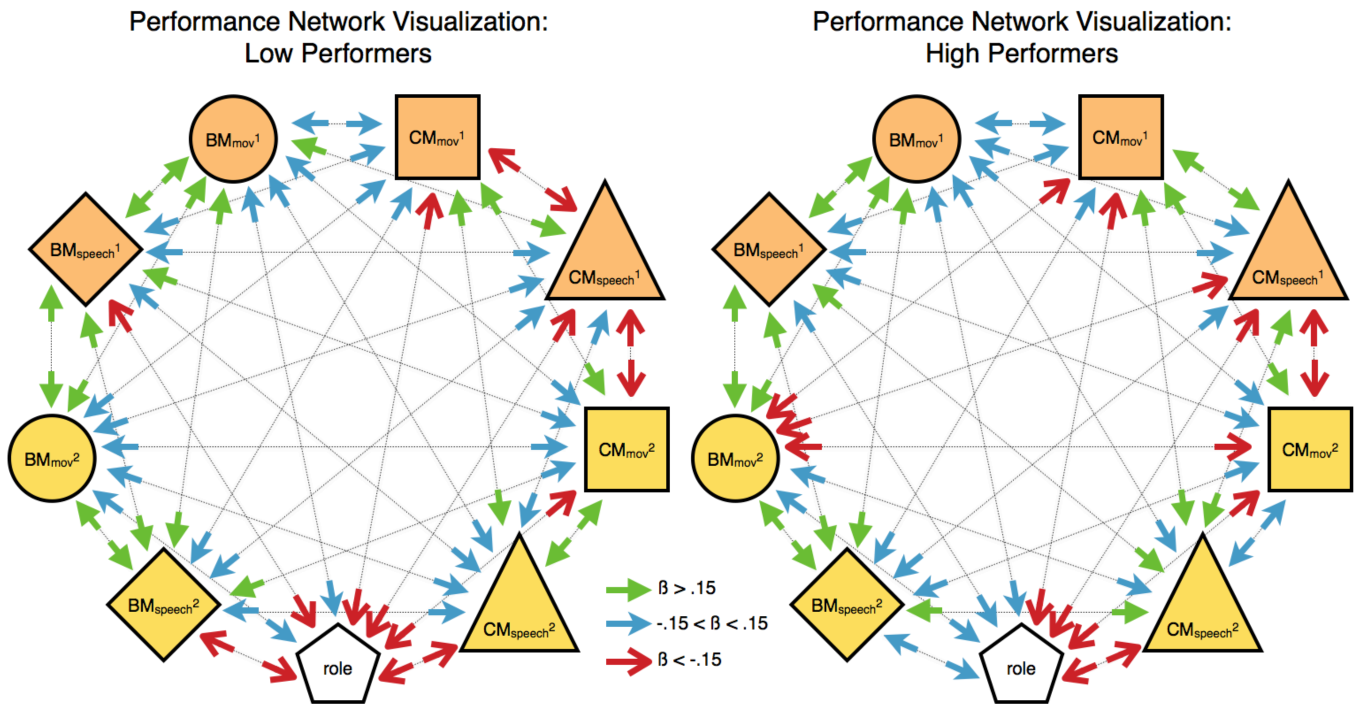

Figure 1: Network visualizations for the low- (left) and high-performing (right) dyads. Connections are effect sizes obtained from linear models between nodes and are color-coded by strength. Nodes represent behavior matching (BM) and complexity matching (CM) for speech and movement (subscript) and half (superscript) and perception of role distribution. Arrows signal bidirectional correlational relationship (not necessarily causal influence) as a graphical convenience.

Performance Network We created two independent networks by grouping the dyads with a median split on the performance variable (see Figure 1). To explore differences in the unfolding of the interaction, both BM and CM measures for the first and second halves of the interaction (noted in superscript) were included as nodes in these networks. Consistent with our hypothesis, high-performing dyads (M = .20, SE = .19) had a lower network strength relative to low-performing dyads (M = .26, SE = .26), $t(71)$ = 3.35, p = .001, suggesting that more flexibly coupled dyadic networks may be free to respond more effectively by converging only in key channels (cf. Fusaroli et al., 2012).

Discussion

Utilizing network visualization techniques (Paxton & Dale, 2013b), the present study provides a first look at the connections across multiple types of interpersonal convergence in multimodal communication. We have presented a network diagram detailing the relationships between behavior matching and complexity matching of movement and speech modalities. By partitioning dyadic-level networks by task performance, we are able to gain insights into differences between high- and low-performing systems. We find that high-performing dyads have statistically lower network strengths than do their low-performing counterparts. This may mean that high-performing dyads have more open degrees of freedom, yielding flexibility that the dyad can leverage to optimize performance on problem-solving tasks.

Additionally, the network analysis structure allows us to qualitatively observe how specific connections change from low- to high-performing dyads. For example, for low-performing dyads, more complexity matching (i.e., lower CM values) of movement during the second half of the interaction predicts less behavior matching, while greater complexity matching in the same setting is associated with more behavior matching in high-performing dyads. Thus, the coordination patterns across multiple time scales changes depending on the performance of dyads. This might relate to a functional mechanism (cf. Louwerse et al., 2012) for multiscale coordination: Higher correspondence of multiple coordination patterns relates to increased communicative benefit (i.e., task performance).

When behavioral synchrony and complexity matching metrics are partitioned across time and modality, differences between the two types of convergence emerge. For instance, behavior matching and complexity matching differentially predict relative dominance in high-performing dyads. Increased complexity matching in both modalities across the interaction strongly predicts participants' beliefs in distinct social roles during the task, whereas behavior matching does not strongly affect the relative dominance construct. We believe this result highlights the importance of studying behavior matching and complexity matching together: While both capture the degree to which individuals affect one another during interaction, each may provide unique insights into patterns of interaction to which the other is blind.

Conclusion

Previous research has supported the existence of behavior matching and complexity matching separately during interaction, but this is (to the authors' knowledge) the first study to examine the two in concert. We have sought to combine the meaningful individual contributions of each level of interpersonal convergence to more fully understand the structure of multimodal communication on surface and statistical levels. Consistent with the view of interaction as interpersonal synergy rather than strict convergence (e.g., Fusaroli et al., 2012; Riley et al., 2011), the present study finds that task performance differs with the interpersonal structure and that optimal performance may be characterized by greater flexibility within the structure. By presenting and analyzing multiscale and multimodal datasets through network visualizations, we have been able to allow the data to suggest interesting future directions for the dataset during our initial investigation of theoretically driven questions.

Acknowledgments

Special thanks go to undergraduate research assistants Jamie Faria, Evelyn Alvarez, and Breanna Wright (University of California, Merced) for their assistance in data collection.

References

Abney, D.H., Paxton, A., Kello, C.T., & Dale, R. (under revision). Complexity matching in dyadic interaction.

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

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(6), 1482–1493.

Clark, H. H (1996). *Using language*. Cambridge University Press.

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

Fusaroli, R., Bahrami, B., Olsen, K., Roepstorff, A., Rees, G., Frith, C., & Tylén, K. (2012). Coming to terms quantifying the benefits of linguistic coordination. *Psychological Science*, *23*(8), 931-939.

Hatfield, E., Cacioppo, J. T., & Rapson, R. L. (2005). Emotional contagion. *Current Directions in Psychological Science*, *2*(3), 96–99.

Helm, J. L., Sbarra, D., & Ferrer, E. (2012). Assessing cross-partner associations in physiological responses via coupled oscillator models. *Emotion*, *12*(4), 748.

Keith, T. Z. (2005). *Multiple regression and beyond*. Boston: Pearson Education.

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

Marmelat, V. & Delignières, D. (2012). Strong anticipation: complexity matching in interpersonal coordination. *Experimental Brain Research,* 222, 137-148.

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

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

Paxton, A., & Dale, R. (2013b). Multimodal networks for interpersonal interaction and conversational contexts. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Meeting of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Garrod, S., & Pickering, M. J. (2004). Why is conversation so easy? *Trends in Cognitive Sciences*, *8*(1), 8–11.

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.

Valdesolo, P., Ouyang, J., & DeSteno, D. (2010). The rhythm of joint action: Synchrony promotes cooperative ability. *Journal of Experimental Social Psychology*, *46*(4), 693–695.

Wujec, T. (2010, February). Tom Wujec: Build a tower, build a team [Video file]. Retrieved from http://www.ted.com/talks/tom_wujec_build_a_tower.html