Quantifying Joint Activities using Cross-Recurrence Block Representation

Tian Xu (txu@indiana.edu)
Computer Science and Cognitive Science Department, 1101 East 10th Street Indiana University
Bloomington, IN, 47405
Chen Yu (chenyu@indiana.edu)
Psychological & Brain Sciences, Cognitive Sciences, and informatics Department, 1101 East 10th Street Indiana University
Bloomington, IN, 47405

Abstract

Humans, as social beings, are capable of employing various behavioral cues, such as gaze, speech, manual action, and body posture, in everyday communication. However, to extract fine-grained interaction patterns in social contexts has been presented with methodological challenges. Cross-Recurrence Plot Quantification Analysis (CRQA) is an analysis method invented in theoretical physics and recently applied to cognitive science to study interpersonal coordination. In this paper, we extend this approach to analyzing joint activities in child-parent interaction. We define a new representation as Cross Recurrence Block based on CRQA. With this representation, we are able to capture interpersonal dynamics from more than two behavioral streams in one Cross Recurrence Plot and derive a suite of measures to quantify detailed characteristics of coordination. Using a dataset collected from a child-parent interaction study, we show that these quantitative measures of joint activities reveal developmental changes in coordinative behavioral patterns between children and parents.

Keywords: cross-recurrence quantification analysis; child parent interaction; multimodal behavioral data; interpersonal interaction; hand coordination; statistical methods

Introduction

Humans are social animals and coordination is the foundation of everyday social interactions (Kendon, 1970). We exhibit a remarkable ability to coordinate our behaviors with our social partners to achieve common goals at different cognitive levels and time scales, from motor physiological coordination, such as dancing (Kimmel, 2012), to goal-oriented dialogue and collective task completion (Fusaroli and Tylén, 2015). During social interaction, the dynamics between interactors evolves and escalates over time in a complex way. To understand how information is exchanged and communication is unfolded through behavioral cues, we need to examine the characteristics of coordination in great details.

Recently, technological advances in sensing and computing devices allow us to collect high-density and large-volume behavioral data to study interpersonal coordination (Dale et al., 2011). However, due to the complexities and stochastic nature of human behaviors, new ways in data collection have also presented data-mining challenges to cognitive scientists (Fusaroli et al., 2014) – with a large amount of data collected, how we can effectively discover novel and reliable patterns to advance our understanding of human coordinated behaviors. For example, in a dyadic interaction, two social partners most often communicate via several expressive channels, such as language, facial expression, body movements, and manual actions. Those intrapersonal and interpersonal behaviors happen in fractions of a second, and together form a dynamically interactive loop to continuously influence each other at every moment (Louwerse et al., 2012).

The primary goal of the present study is to introduce a new method to quantify engagement and joint activities in social interaction. The method is built upon Cross Recurrence Plot Quantification Analysis (CRQA), which has been introduced in cognitive science to analyze synchronization and alignment in interpersonal coordination (Riley et al., 2011). The proposed method extends classic CRQA approaches by using a new representation of joint events: Cross Recurrence Block. Multiple behavioral measures can be derived from this new representation, revealing fine-grained quantitative patterns that cannot be detected with standard statistical measures of individual and joint behaviors.

Cross Recurrence Plot Analysis

Recurrence plot Quantification Analysis (RQA) and Cross-Recurrence plot based Quantification Analysis (CRQA) have been proposed as nonlinear methods that provide informative data visualization and rigorous quantification of dynamical systems and their trajectories with which many properties would otherwise be lost due to averaging with traditional correlation analysis (Zbilut et al., 1998; Marwan et al., 2007). In particular, CRQA is able to reveal and describe the shared dynamic trajectory between two different data series by constructing the states that both systems visit over time. While first introduced as a form of generalized cross-correlation between two continuous data streams, CRQA has been extended to nominal time series with categorical values using a radius of zero in computing state matches between two systems (Coco and Dale, 2014). These methods have been widely used in many areas, such as the analysis of pattern and rhythm in sound and music (Cooper and Foote, 2002), visual search scan pattern analysis (Anderson et al., 2012) and discourse conceptual structure analysis in doctor-patient conversations (Angus et al., 2012). Despite the analytical power of CRQA as a non-linear method for extracting inter-system dynamics, the process of constructing Cross Recurrence Plot (CRP) can be relatively simple and straightforward. In the next section, we will introduce the plot

construction process and CRQA with the empirical dataset from a parent-child toy play interaction study.

Generating Cross-Recurrence Plots from Data

Figure 1 shows both the experimental setup and multimodal data streams collected in our study. We invited 24 dyads of children and their parents to the lab (12 children at 12 months old age, 12 children at 24 months old age). During the experiment, parents and children played with two sets of toys in free-flowing interaction for about 6 minutes (Yu, Smith, 2013). The toys were painted into single colors (blue, green or red) to facilitate automatic image processing. We recorded the entire interaction from multiple cameras, including both the child's and parent's first-person viewpoints, a third-person view, and a bird-eye view. The present study focuses on manual activities generated by parents and children in toy play. Human coders went through the videos from multiple viewpoints and manually annotated frame-by-frame about which object was held by the child and the parent with each of their hands. As shown in Figure 1(b), holding actions of both the child and parent were represented as four categorical temporal data streams with different color indicating different Region-Of-Interest (ROI) in each moment. The
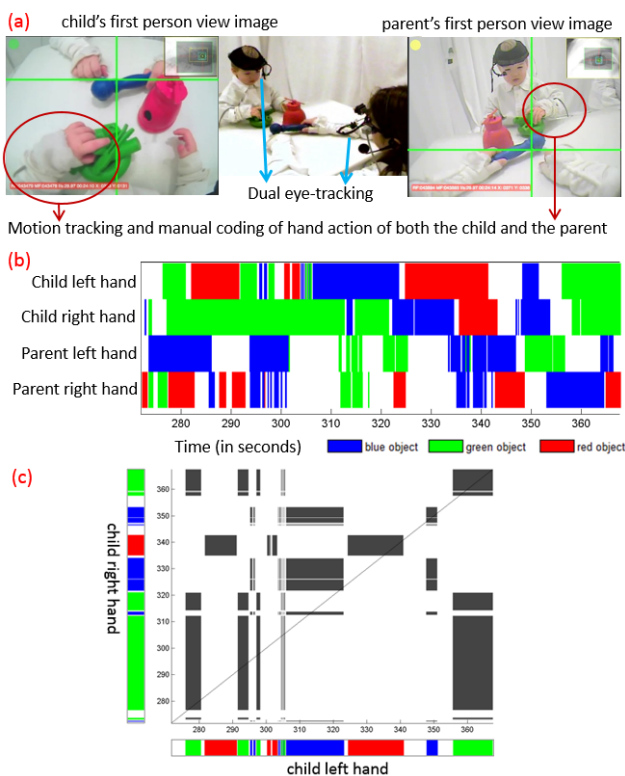

Figure 1: (a) Our multisensory experimental set-up wherein the child and parent engaged in free object-play interaction. Both child's and parent's gaze, holding action and speech events were coded into nominal temporal data streams; (b) Coded manual holding action behavioral streams during one experimental trial; (c) A Cross Recurrence Plots constructed using a child's two hand action streams with the data shown in (b).

sampling rate of all four time series was 30Hz. Each data point in a time series has been assigned with one of the four possible ROI values: 1-3 (blue, green, red) indicating the target object held by one of the interactors, and 0 (white) meaning empty hand.

Figure 1(c) shows a Cross Recurrence Plot (CRP) constructed with the child's left hand holding data stream on the x-axis and the child's right hand holding data stream on y-axis. A CRP is basically a M×N matrix, in which N is the number of data points of the temporal stream on x-axis and M is the number of data points on y-axis. For any cell in the matrix at row i column j, that point $R_{i,j}$ in CRP is assigned to 1 if the child's right hand at time t_i and the child's left hand at time t_j were both holding the same object; $R_{i,j}$ will be 0 otherwise (Marwan et al., 2007).

With standard CRQA, we can calculate various measurements such as recurrence rate, determinism, etc. (Zbilut et al., 1998). For example, recurrence rate calculates the density of recurrence plots that reflects the degree of coordination and coupling between the behavioral streams on x-axis and y-axis. In Richardson and Dale (2005), the authors calculated the recurrence rate within a 30-second temporal window to reflect the coordinated gaze behaviors between the speaker and listener; In Yu and Smith (2013), the authors calculated correlation scores with different time delays and recurrence rates within multiple temporal windows to reveal how eye-hand coordination between child and parent influences the child's visual attention towards objects.

Cross Recurrence Block Representation

In standard CRP, $R_{i,j}$ can be either 1 (the behavior of the interactor on x-axis at time t_i and his partner on y-axis at time t_j fall in the same category), or 0 (no match in their behaviors). In such cases, a plot is usually visualized using two colors, black and white (Marwan et al., 2007). The first contribution of our new method here is to extend the value range of $R_{i,j}$ to include the different behavioral categories; and as a result, the CRP becomes colorful – each color depicts not only a match between two data points from two time series at a certain point in time but also the exact behavioral category that is matched. Figure 2 shows the colored version of CRP constructed with child's holding data shown in Figure 1(c). In this plot, we used held objects as behavioral categories.

Traditionally, the data structure of Cross Recurrence Plot is represented as a matrix composed of individual points and continuous lines. Therefore, quantitative measurements derived are based on those two representations. As illustrated in Figure 2, a set of temporally adjacent recurrence points form a block (labeled as $Block_A$ with green color). When we look back to the original data streams recording the held objects in the child's two hands (shown at the bottom of Figure 2), one can easily see that $Block_A$ is the product of a sequence of two individual holding events: the child was holding the green object from 276s to 280.5s in his left hand, and his right hand also grabbed the same object at 276.8s and continued to hold it much longer until 312.2s. One can
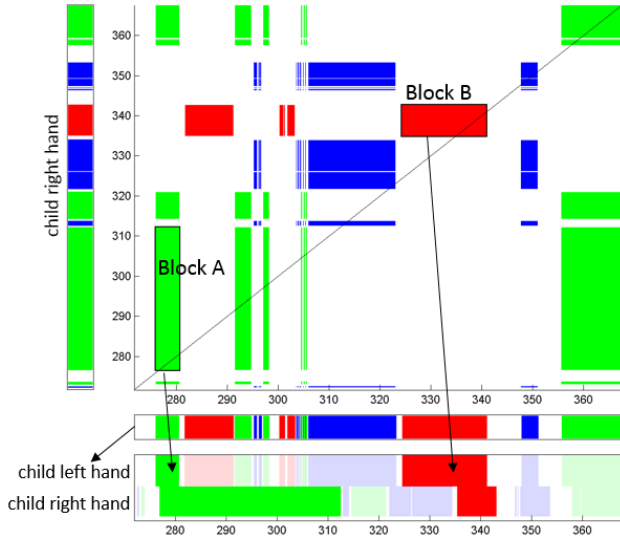
Figure 2: Cross Recurrence Plot with added color information indicating the target object of each joint activity sequence.

intuitively see that $Block_A$ captures a sequential pattern of two manual actions. Similarly, $Block_B$ represents another pair of sequential holding events: the child was playing with the red object using their left hand first from 324.3s to 340.8s, then their right hand let go of the blue object to jointly hold the same red object from 335.1s to 343.6s.

As illustrated by the two examples above, we propose a new representation based on CRP, viewing and grouping temporally adjacent points as blocks with width and height. Such blocks represent joint activities that can be directly mapped back to temporal events in the behavioral data streams used to construct a cross recurrence plot. Formally, we define such blocks as **Cross Recurrence Block (CRB)** that are made of a set of points in a CRP sharing the same categorical value and being temporally adjacent to each other. As shown in Figure 2, a CRB can vary in size and shape; if the height of one CRB is just one data point, this CRB is equivalent to a horizontal line; if, in an extreme case, both the height and width of one CRB is equal to one data point, this CRB is in fact just one single point. With this general representation, the same criterion and measurements can be applied to all CRBs, in contrast to point and line representations which always require different sets of parameters.

In implementation, every CRB can be simply represented as a vector $<x_{min}, y_{min}, x_{max}, y_{max}, cv>$ describing the location of this CRB on the cross recurrence plot and its categorical value. Thus, any recurrence matrix/plot can be economically stored as a list of CRBs: $\{CRB_1, CRB_2 \dots CRB_x\}$. This method dramatically reduces both storage and computation costs, compared with the traditional way of saving the entire M×N matrix.

To summarize, the new representation CRB has three advantages: 1) it contains the information about categorical values in two data streams, not just whether it is a match or not; 2) it can be directly linked to joint action sequences in

behavioral data; and 3) it is an efficient data structure to represent and store the entire plot for any further computation.

CRP with more than two data streams

An additional advantage of CRB is to capture complex joint activities. Due to the data structure of a single matrix used in standard CRQA, each plot can only be constructed to reflect the interaction dynamics between two data streams (Marwan et al., 2007; Fusaroli et al., 2014). With CRB representation, blocks can overlap with each other. Therefore, multiple data streams can be used in either x-axis or y-axis. Figure 3(a) shows a CRP constructed with two holding action streams from child on the x-axis and two holding action streams from parent on the y-axis. For any point at (row_i, $column_j$) in the current CRP, we compare the i_{th} data points in all the data streams on y-axis at time t_i with all the j_{th} data points in temporal streams on x-axis at t_j. The categorical value of that point can be a list if there are more than one match.

$$R_{i,j} = \{cv_{i_1}, cv_{i_2}, \dots, cv_{i_k}\} \cap \{cv_{j_1}, cv_{j_2}, \dots, cv_{j_k}\} \qquad (*)$$

cv_{i1} is the categorical value of i_{th} data point in the first data stream on y-axis; cv_{j1} is the categorical value of j_{th} data point in the first data stream on x-axis.

For example, point p' in Figure 3(a) has two categorical value matches (blue and green) because the child and parent were engaged in playing both the green and blue objects intermittently. More specifically, such joint play activities were usually formed because both partners played with one object at the beginning, then jointly switched to another object, and then jointly switched back to the one they played previously. In this case, the engagement between the first period and the third period on the original object form a large cross recurrent block which is overlapped with the cross recurrent block of the second in-between object. In implementation, after obtaining the two matching categorical values of p', we group them into the existing CRBs. In this example, point p' belongs to two blocks: a part of green CRB_1, and also a part of blue CRB_2. In this way, point p' contributed to the overlapping portion between CRB_1 and CRB_2. The overlapping situation between two CRBs with multiple categorical value assignments of point p' is illustrated in Figure 3(b) with each layer consisting of one CRP with one different categorical value.

As shown in Figure 3(b), the recurrence matrix could also be stored in 3 dimensions – one for each categorical value. However, in practice, using M×N×C as the recurrence data structure (C is the total number of categorical values in behavioral data streams) would require a large storage space. In addition, this 3D matrix can be sparse, and therefore both storage and computing costs can be high when operated on a 3D sparse matrix representation. Instead, using CRBs as the basic building component to represent a cross recurrence plot has two advantages: 1) **completeness**: it allows multiple categorical value assignments of any point in a recurrence matrix because it can handle overlaps between different
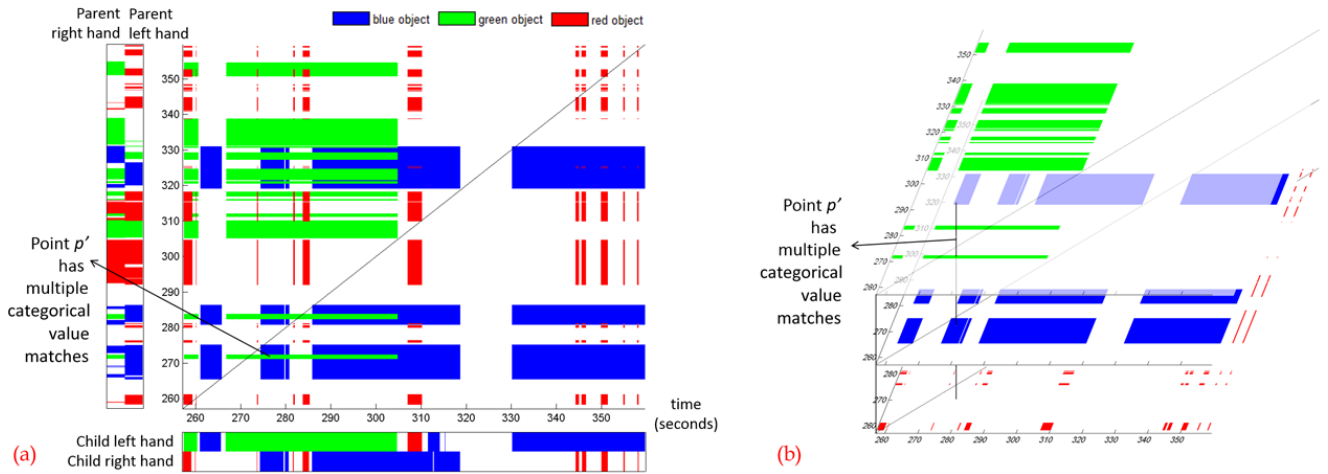
Figure 3: (a) A Cross Recurrence Plot constructed with two hand action streams from the child on the x-axis and two hand action streams from the parent on the y-axis. **C**ross **R**ecurrence **B**lock (CRB) representation allows a single plot to capture the dynamics from more than two behavioral modules. (b) Point *p'* has two matching categorical values and constitute one overlap point between two CRBs.

blocks, and 2) **efficiency**: it avoids unnecessary storage and computing costs.

With our new representation, Cross Recurrence Block, a variety of measurements can be extracted to reflect detailed interaction patterns. Next, we will introduce a suite of quantitative engagement measurements derived from CRBs.

Coordination beyond Synchronization

Interpersonal coordination is conducted in many complex forms that goes beyond synchronized behaviors wherein two partners were engaged in the same object/task at the exactly same time. In daily conversations and joint activities, we lead and follow each other's behavioral cues, take turns, diverge and converge from one target onto another. In this section, we will show that the CRB representation can be used to reveal fine-grained patterns of real-time coordination.

Width and Height

For each block, the width equates the horizontal length of each CRB and reflects how long the behavioral module(s) of $agent_x$ on the x-axis has participated in this joint action episode. Similarly, the height measure states how long the behavioral module(s) of $agent_y$ on the y-axis contributed to the same joint activity.

For example, in Figure 4, $Block_B$ is formed by the child's holding the blue object from 285.9s to 318.6s and then the parent's holding the same object from 280.5s to 286.3s. Thus, $Block_B$'s width is 32.7s and its height is 5.8s. Similarly, the sequence of holding events that formed $Block_C$ is that the child was holding the red object first from 307.1s to 310.2s, then the parent picked up the same red object and held it from 292.1s to 304.6s. As a result, $Block_C$'s width is 3.1s and its height is 12.5s.

Shape

For every joint action, if the child was holding the object

longer than the parent, then the formed CRB appears wider, as horizontally shaped, such as $Block_B$ in Figure 4; instead, if the parent was holding the object longer, the corresponding CRB will be more vertically shaped, such as $Block_A$ and $Block_C$. The frequencies of horizontally and vertically shaped CRBs show the overall influence of each agent in joint activities throughout the whole interaction.

Time lag

The relative position of each CRB respect to the diagonal line reflects the leading and following relationship between the two participants. For each CRB, the lower left corner indicates the starting point of this joint activity sequence and the upper right corner reflects the ending point where both agents exit the joint action episode one after another (see $Block_C$ in Figure 4). Thus, the start time lag is calculated by subtracting $time_x$ from $time_y$ of the lower left corner of the
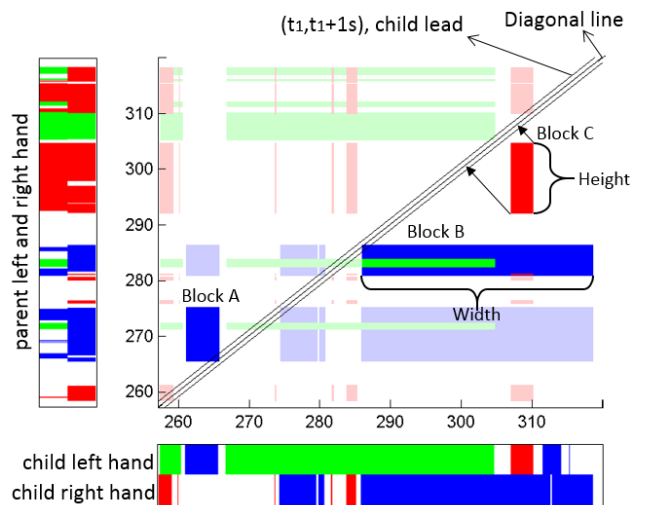

Figure 4: Illustration of Cross Recurrence Block based quantitative measurements

block: $start\_time\_lag = time_y - time_x$. If the time lag is positive, then it means that for this action sequence, $agent_x$ acts earlier than $agent_y$, who is the initiator of this activity episode. For the end time lag, similarly we calculate it by subtracting $time_x$ from $time_y$ of the CRB's upper right corner. With visualization, one can see the leading and following relationship simply by judging the positions of the start and ending points relative to the diagonal line in the plot.

With highlighted synchronized diagonal line and time shifted diagonal lines in Figure 4, we can intuitively identify the leading and following agent of each joint activity episode. For example, in $Block_A$, most part of which is above the diagonal line, was formed by the sequence of events that the child held the blue object first, then passed the object onto the parent's hand. In contrast, in $Block_B$, the parent was playing with the blue object first at 280.8s, then the child joined the object play at 285.8s, and continued playing with the same object until 318s, long after the parent switched to play with other objects at 286.3s. The start time lag of $Block_B$ is -5.0s, meaning the parent's action preceded the child's by 5 seconds; and the end time lag is -31.7s, meaning the parent switched to other objects first.

In addition, the numerical value of this measure and CRB's relative position to the diagonal line directly display the temporal relevancy of the event sequence that constitutes each individual CRB. The CRBs in upper left and lower right corners of the plot are formed by events far away in time. In this paper, we mainly focused on CRBs that overlapped with the diagonal line and calculated their CRB based quantitative measurements, which will be shown in the next section.

Results

With the holding action dataset collected from our child-parent toy play study in two different age groups (12 and 24 month), our main goal here is to demonstrate that standard measures of individual and joint behaviors didn't reflect different coordinative patterns between the two age groups, while the measures based on CRB representation revealed fine-grained patterns showing different interaction dynamics in child-parent play.

To compare with CRB-based measures, we first computed standard behavioral statistics, such as the mean duration and frequency of the child's and parent's holding events, mean duration and proportion of time of joint holding events wherein the child and the parent were holding the same object with either or both of their hands. The mean results and standard errors are listed in Table 1. None of those measures is statistically significant.

Next, we applied **C**ross **R**ecurrence **B**lock based **Q**uantification **A**nalysis (CRBQA) to the dataset, constructed CRP with four holding action data streams for each individual trial based on CRB data structure, and calculated quantitative measures including width, height, start and ending time differences, frequency of vertical and horizontally shaped CRBs overlapped with the diagonal line.

Table 1. List of behavioral measures of the child's and parent's holding events between two age groups

Behavioral measure	12 month	24 month	stats
Duration of child holding (seconds)	4.67±0.51	3.42±0.43	t(23)=1.87 p=0.07
Duration of parent holding (seconds)	1.91±0.14	2.40±0.21	t(23)=-1.90 p=0.07
Frequency of child holding (per minute)	15.49±1.98	15.26±1.30	t(23)=0.09 p=0.93
Frequency of parent holding (per minute)	22.42±1.56	22.39±1.90	t(23)=0.01 p=0.99
Duration of joint holding (seconds)	0.82±0.07	10.71±0.05	t(23)=1.23 p=0.23
Joint holding Proportion of time	11.27±1.65	11.24±1.22	t(23)=0.01 p=0.99

The CRB based results are shown in Table 2. Width indicates the duration of child's participation in forming CRBs, while height states for the parent's holding duration. Horizontal CRBs are blocks wherein the child's holding lasted longer than the parent, and vice versa for the vertical CRBs. With start time lag, positive values indicate that the joint holding sequences were initiated by the child and how many seconds afterwards the parent followed up by holding the same object; for ending time difference, positive values show that the child stopped holding the object first and negative values show that the parent switched to a new target object first when the child was still holding the same object.

Table 2. List of measures extracted with Cross Recurrence Block based Quantification Analysis of the child's and parent's holding events between two age groups

CRBQA	12 month	24 month	stats
Width (seconds)	11.89±2.34	7.96±1.45	t(23)=1.43 p=0.17
Height (seconds)	3.25±0.30	7.33±1.43	t(23)=-2.80 p<0.01
Frequency of horizontal CRBs (per minute)	4.51±0.56	3.82±0.33	t(23)=1.05 p=0.30
Frequency of vertical CRBs (per minute)	2.22±0.54	4.57±0.91	t(23)=-2.23 p<0.05
Start time lag (in seconds)	4.55±1.75	-0.01±1.32	t(23)=2.08 p<0.05
End time lag (in seconds)	-4.09±0.93	-0.63±1.26	t(23)=-2.21 p<0.05

Among 6 measures based on CRB representation, four of them show significant differences between the two groups. 12-month-old infants seem to hold objects much longer in joint activities compared with their parents (11.89 vs. 3.25 seconds). At the 24 month, they seemed to more or less become equal partners (7.96 vs. 7.33 seconds). For the measure of start time lag, positive values indicate that the joint holding sequences were initiated by the child. In the 12-month-old group, after the child started playing with a target object with their hands, on average, 4.55 seconds later, the parent followed up by holding the same object. For end time lag, the negative values for 12 month old group meant that the parent switched to a new target object first 4.09 seconds on average before the child terminated his manual actions on the current object. For the 24-month-old group, there was not much difference on temporal leading and following relationships between the dyad. This suggested that in the 12-month-old group, the parent was mainly following the child's lead, and in most joint plays, the child started holding a target object first and continued to play with it even after the parent switched to hold another object. These examples here show promise that the CRB-based approach can reveal fine-grained patterns of joint activities that we may not be able to obtain otherwise.

Conclusion

In this paper, we introduced a novel method to analyze coordinated behaviors between two social partners. This approach is built upon Cross Recurrence Quantification Analysis. The three key contributions here are: 1) to encode state matches between two agents in different categorical values and colors in a Cross Recurrence Plot (CRP); 2) to use a block representation to capture interaction with more than two data streams in one plot; 3) to derive a suite of measures based on CRB representation. We used holding action dataset from a child-parent interaction toy play study to demonstrate that this method can capture characteristics and dynamics beyond two temporal streams, revealing developmental changes of coordinated behaviors between 12 month and 24 month. One future direction in this particular application is to show coordinated patterns extracted based on CRB representation play a critical role in child development. For example, if better coordination leads to better development, then we should be able to use some derived measures to predict later development outcome. Moreover, this method can be applied to basically all interaction contexts with categorical behavioral data on various time scales. Any form of interpersonal or intrapersonal engagement shared between different behavioral models can be easily visualized and quantified with our method, such as gaze, speech data, pointing gestures, etc. As we move toward data-intensive science to understand human cognition, we will rely more heavily on developing and utilizing data analysis methods, such as the one presented here, to look for findings that may be undiscovered but highly robust. We envision that this method to be a useful and intuitive analytical tool for interaction study community.

References

Anderson, N. C., Bischof, W. F., Laidlaw, K. E., Risko, E. F., & Kingstone, A. (2013). Recurrence quantification analysis of eye movements. *Behavior research methods*, *45*(3), 842-856.

Angus, D., Smith, A., & Wiles, J. (2012). Conceptual recurrence plots: Revealing patterns in human discourse. *Visualization and Computer Graphics, IEEE Transactions on*, *18*(6), 988-997.

Coco, M. I., & Dale, R. (2014). Cross-Recurrence Quantification Analysis of Behavioral Streams: An R Package. *Frontiers in Psychology*, *5*.

Cooper, M. L., & Foote, J. (2002). Automatic Music Summarization via Similarity Analysis. In *ISMIR*.

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.

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

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

Kendon, A. (1970). Movement coordination in social interaction: Some examples described. *Acta psychologica*, *32*, 101-125.

Kimmel, M. (2009). Intersubjectivity at close quarters: how dancers of Tango Argentino use imagery for interaction and improvisation. *Cognitive Semiotics*, *4*(1), 76-124.

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

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

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.

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

Yu, C., & Smith, L. B. (2012). Embodied attention and word learning by toddlers. *Cognition*, *125*(2), 244-262.

Yu, C., & Smith, L. B. (2013). Joint attention without gaze following: Human infants and their parents coordinate visual attention to objects through eye-hand coordination. *PloS one*, *8*(11).

Zbilut, J. P., Giuliani, A., & Webber, C. L. (1998). Detecting deterministic signals in exceptionally noisy environments using cross-recurrence quantification. *Physics Letters A*, *246*(1), 122-128.