From Data Streams to Information Flow: Information Exchange in Child-Parent Interaction

Heeyoul Choi, Chen Yu, Linda B. Smith, and Olaf Sporns (chenyu@indiana.edu)
Psychological and Brain Sciences, and Cognitive Science Program
Indiana University, 1101 E. 10th St., Bloomington, IN 47405

Abstract

The goal of this paper is to enhance understanding of how bodily actions between two social partners are coordinated in interpersonal interactions in naturalistic contexts. To this end, we introduce information-theoretic measures as a new approach to capturing sensorimotor dynamics in child-parent social interaction. In particular, information flows were measured based on a set of variables extracted from multimodal fine-grained behavioral data in social interactions wherein a child and a parent played with a set of novel toys. Our results showed that information-theoretic measures can indeed capture the inherent structure of perception and action dynamics and further information exchange patterns can be used to predict successful learning through child-parent interactions. Moreover, those information flows between sensorimotor variables reveal a set of underlying perceptual and motor patterns with cognitively plausible explanations. In summary, the present study represents the first steps to connect information-theoretic measures as a mathematically rigorous framework with embodied human communication and cognition.

Keywords: Embodied Cognition; Local Transfer Entropy; Information Theory; Word Learning; Social Interaction.

Introduction

How do two interacting agents couple their activity? Some forms of human collaborative and coordinated behavior (such as maintaining a conversation, or jointly solving a complex problem) appear to happen effortlessly as if the participants can read each other's mind and understand each other's communicative intent. At an elementary level, inter-agent coordination depends on external (and observable) behaviors by the participants where the behavior of one participant influences the behavior of the other. Past research tells us that behaviors such as eye movements, head turns, and hand gestures are critical to this coordination. However, very little is known about the *real-time* dynamics of these behaviors in social interactions nor about how they may be related to higher-order functions such as making inferences about the goals and intentions of other.

Because so little is known about the real time dynamics of the sensory-motor behaviors on which social coordination rests, the present study takes a bottom-up approach, measuring multiple sensory streams – head and hand movements as well as each participant's view of the events – and then attempts to determine the possible signatures of coordination in these behaviors. The social coupling of a toddler and a parent is an appropriate first setting for this endeavor because the toddler as a developing system is just learning the relevant sensorimotor cues and thus may enable us to see more clearly the strands that are more tightly woven in the highly developed adult system.

The key idea of the present study is to apply information-theoretic measures to understand the structure in the sensorimotor dynamics of the interaction. To this end, we conceptualize multimodal information flows between children and parents as those between senders and receivers in artificial communication systems (Shannon, 1948). More specifically, the child and the parent communicate with each other using multiple communication channels such as gaze, pointing, speech, and hand movements. The specific goal this study is to understand how information theoretic measures might be used to analyze the flow and information exchange within each participant and between participants. For example, within an individual, do behaviors such as looking "send" information to the hands, in the sense of signalling a reach? Across individuals, does a hand action by one participant send information to the gaze of the other? And, if we can measure information flow in these ways, can we also measure how it might change at different points in the interaction, for example, when an object is being named?

Historically, information theory was developed to find fundamental limits on compressing and reliably communicating data within single transmission channels (Shannon, 1948). Since its inception, information theory has found applications in many other areas, including statistical inference, natural language processing, the evolution and function of molecular codes, model selection in ecology, thermal physics and other forms of data analysis(de Ruyter van Steveninck et al., 1997). Recently, information theory has been applied in the context of embodied autonomous systems to help characterize the flow of information between (neural or algorithmic) control architectures, body and environment (Sporns & Lungarella, 2006).

Despite the recent success of information-theoretic measures in various scientific fields, these recent advances have not been systematically applied to human behavioral data. Thus the present study seeks both to understand the sensorimotor dynamics of social interactions as information flow and to develop a mathematically rigorous framework within which to do so.

Experiment and Data Preprocessing

Figure 1 provides an overview of the approach. We measured multiple sensory streams with no prior expectations that they are independent or dependent. These
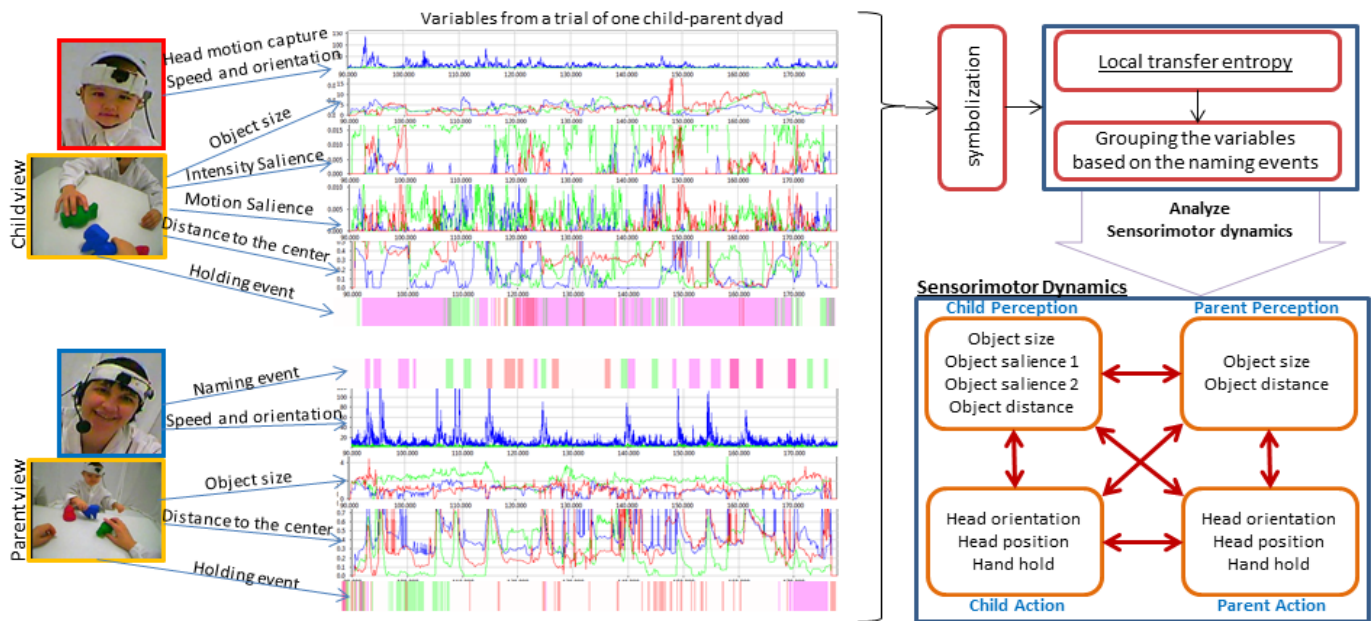
Figure 1.Overview of our approach. Sensorimotor dynamics are analyzed based on the normalized transfer entropy around naming moments. 1) **Data**: Multimodal time series were collected from child-parent interactions. 2) **Symbolization and information-theoretic measure**: those continuous time series were converted into symbolic sequences and information flows between those sequences were calculated. 3) **Pattern Discovery**: finally we analyzed information transfer patterns between the child and the parent and discovered psychologically interesting patterns that can lead to smooth social interaction and learning.

continuous data streams are then subjected to a *symbolization step* that, in a mathematically defensible way, partitions the continuous values into a set of discrete categories. From these, we can further apply information-theoretic measures by first grouping the temporal variables according to different categorical events in the interaction, and then measuring the information flow between and within participants.

Before we provide further details, we briefly review the experimental setup, the nature of the multimodal data and data processing. More details can be found in Yu, Smith, Shen, Peirera, & Smith (2009) and Yu & Smith (under revision). In the experiment, the child and the parent sat opposite each other at a small table and the parent was instructed to interact naturally with the child, engaging the child's attention with the toys while naming them. The toys and names were novel to the children. In total there were six objects with novel shapes and solid colors and 6 to-be-learned object names were artificial names (e.g. "bosa", "dodi"). Children and parents played with three objects at a time. Eight children between 18 to 24 months of age and their parents participated in the study.

There were two sessions of the study. In the free-play session, parents and children played with 3 objects for 4 1.5 minute play periods. Parents were asked to play naturally and if they named the objects, they had to use the names that were provided by the experimenter and pre-taught to the parents. The novel aspect of the study was the multimodal sensing equipment worn by the participants: two head-mounted mini cameras that were placed on both the child's and the parent's foreheads, motion tracking sensors placed

on heads, and audio recording of the parent's speech. The head cameras captured the dynamic visual information from each participant's first-person perspective.

After these play trials, the child was tested to determine whether the child had learned any of the object names. This name-comprehension test was performed by the experimenter. On each trial, three objects were placed out of reach of the child about 30 inches apart, one to the left of the child, one in the middle and one to the right. Then the experimenter looked directly into the child's eyes, said the name of one of the objects and asked for it. Direction of the child's eye gaze was scored as indicating comprehension. Each word was tested twice with a score ranging from 0, 1 to 2. The objects with score 0 or 2 are considered to be unsuccessfully or successfully learned respectively.

Data Preprocessing

The data from head cameras, motion sensors and audio were automatically annotated by various image and sensory processing tools developed in our previous work. Technical detailed can be found in Yu, Smith, Shen, Pereira, and Smith (2009).

Video processing. The recording rate is 30 frames per second and the resolution of each image frame is 720x480. The image data is analyzed in two ways: (1) At the pixel level, we use the saliency map model developed by Itti, Koch, and Niebur (1998)to measure which areas in an image are most salient based on motion and intensity. (2) At the object level, we automatically extract visual information, such as the locations and sizes of objects, from sensory data in each of the two cameras. The combination of using pre-

defined simple visual objects and utilizing state-of-the-art computer vision techniques results in high accuracy in visual data processing.

Motion data processing. Two motion tracking sensors on participants' heads recorded 6 degrees of freedom of their head movements at the frequency of 240 Hz. Raw motion data {x, y, z, h, p, r} from each sensor were grouped into position {x, y, z} and orientation {h, p, r} groups, and then a motion detection program was developed and used to compute the magnitudes of both position movements and orientation movements.

Speech processing. The parent's speech was recorded. The speech signals were processed and we counted a spoken utterance sentence containing an object name as a naming moment for that object.

From Data Streams to Information Flow

We begin with multiple streams of continuous data from sensorimotor child-parent interaction. Yet the goal is to measure information exchange at the bit level. Thus, we need to convert continuous time series into streams of discrete states. From these, we can then form probabilistic distributions of the states of each variable over time, and apply information metrics to quantify the amount of information in bits.

Symbolization

Symbolization is the procedure in which a continuous data stream is converted a symbol sequence. As in the SAX algorithm (Lin, Keogh, Lonardi, & Chiu, 2003), we used the distribution of piecewise aggregate approximates (PAAs) and made a uniformly distributed symbol set based on the histogram of the PAAs. Compared with other approaches, SAX allows lower-bounding distance measures to be defined on the symbolic space that are identical with the original data space. Thus, the information loss through this symbolization and its potential effects on subsequent data processing is minimal. One open issue in this process is window size; it needs to be neither too small nor too large to capture the relevant dynamics of the phenomena under study. Here we set the window size to 300 msec (PAA=3), as most micro-level human behaviors, such as gaze fixations, happen at this timing scale.

Transfer Entropy

With symbolic representations of various derived time series from multimodal child-parent communication, we can select two time series and measure transfer entropy as a way to capture information transfer between the two underlying processes that generate these two time series. Transfer entropy was originally proposed by Marko (1973), and has been successfully applied to understand how much information flows between variables (Massey, 1990;Schreiber, 2000; Gourevitch&Eggermont, 2007). This

section provides some technical details in information theoretic calculations.

For example, in order to measure the potential information transferred by the child's head movement to the child's hand actions, let X be a symbol sequence representing the child's head movement and Y be a symbol sequence representing the child's hand actions, the transfer entropy from Y to X can be calculated through the following steps. First, let A be a set of the symbols used in sequence X, and $p(x)$ be the probability mass function of symbol x. *Entropy* of sequence X is defined by

$$H(X) = -\sum_{x \in A} p(x) \log p(x).$$

Next, we calculate *conditional entropy*, $H(X|X^p)$, as the entropy in sequence X given its previous values of X:

$$H(X|X^p) = -\sum_{x \in A, x^p \in A} p(x, x^p) \log p(x|x^p).$$

$H(X|X^p)$ can be viewed as the amount of information transferred from one's previous state to its current state in the same sequence/process.

Further, given two processes X and Y, *free information* of X given X^p and Y^p is defined by:

$$H(X|X^p, Y^p)$$
$$= -\sum_{x \in A, x^p \in A, y^p \in A} p(x, x^p, y^p) \log p(x|x^p, y^p).$$

Free information captures the amount of information of X that is not transferred from Y^p as Y^p is given (Note in information-theoretic measures, a variable with certainty doesn't contain any information).

Now that given conditional entropy and free information, *transfer entropy* from Y to X is defined by

$$TE(Y, X) = H(X|X^p) - H(X|X^p, Y^p),$$

which is the information in current X coming from Y. Thus, by subtracting free information in X which has nothing to do with Y (as Y^p is given), we can obtain the amount of information actually transferred from Y to X.

Finally, in practice, two additional steps were included to improve calculation accuracy. The first step was to normalize transfer entropy from Y to X with respect to the total information in sequence X itself. In this way, we can obtain the *relative* amount of information transferred by Y. Moreover, we introduce shuffled transfer entropy $TE(Y^s, X)$ as a bias term (Y^s is a shuffled sequence of Y) to remove accidental information captured by the information between X and Y^s, as Y^s contains the same symbols as those in Y but those symbols are arranged in a randomly shuffled order. As a result, the final form used in our study is called *normalized transfer entropy*:

$$NTE(Y, X) = \frac{TE(Y, X) - TE(Y^s, X)}{H(X|X^p)}.$$

Finding Cognitively Significant Events

If these information theoretic measures capture the structure in the parent-child interaction, then they should be revealing

about significant behavioral events in that interaction. The naming of objects is known to a psychologically compelling moment in parent and child interaction and one in which the parent and child seem to tightly coordinate attention (Tomasello & Farrar, 1986). Accordingly, we grouped the normalized transfer entropy sequence into three time periods defined around naming events: 1) "during" moments defined by the onset and offset of a naming event; 2) "before" moments defined by 3 seconds prior to the onset of a naming event to that onset; 3) "after" moments defined by the offset of a naming event to 3 seconds after that offset. In addition, we segregated those naming events into two kinds, those that led to successful learning and those that did not (as measured by the comprehension test after play). Thus, the goal was to compare information transfer patterns in successful learning with those in unsuccessful learning, and at three particular (and critical) moments in the interaction (before/during/after naming).

In the present paper, we selected a subset of sensorimotor variables in the interaction, grouped them into 5 semantic categories and reported information exchange patterns between them: child's perception, child's head movements, child's holding action, parent's head movements and parent's holding action. As shown in Figure 2, four variables from the child's perception (1-4) contain visual information of named objects, such as its size, intensity saliency, motion saliency, and its spatial location in the head camera view. Four variables (17-20) from the child's head movements measures both orientational and positional changes of the child's head. Similarly, four variables from the parent's head movements (21-23) capture the dynamics of the parent's head. Additionally, variables 7 and 8 contain the information on which objects held by either the child or the parent.

The first questions we asked were whether the transfer entropy measures captured the inherent structure of perception versus action and of one participant versus the other, and whether information transfer patterns in successful learning differ with those in unsuccessful learning. To this end, we first calculated all of the transfer entropies between any pairs of two variables in the data set which formed a transfer entropy matrix. Within this matrix, each cell indicates the amount of information transferred from one variable to the other. We viewed this transfer entropy matrix as a similarity matrix and applied multidimensional scaling (MDS) to recover the structure between those variables based on their information exchange. This data analysis procedure consists of two steps. First, a normalized transfer entropy matrix was converted into a symmetric dissimilarity matrix. Next, multidimensional scaling (MDS) equipped with the constant adding technique as in kernel Isomap (Choi & Choi, 2007) was applied to the dissimilarity matrices.

Given we decompose and group temporal sequences into six groups – before/successful, during/successful, after/successful, before/unsuccessful, during/unsuccessful, and after/successful, six normalized transfer entropy

matrices were computed and their MDS plots were shown in Figure 2, which reveals various patterns and dynamics between five variable categories. In the following, we will further quantify those patterns with cognitive interpretations.
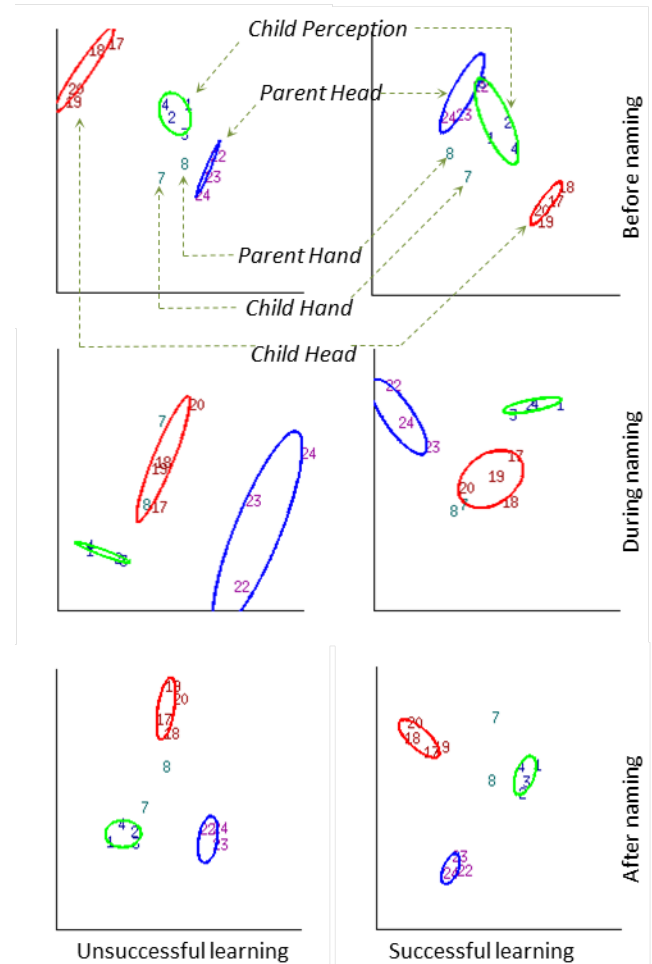

Figure 2. MDS plots from 6 normalized transfer entropy matrices: (Left column) unsuccessful learning, (Right column) successful learning. (Top) before the naming moments, (Middle) during the moments and (Bottom) after the moments. The red ellipses are for child's head movement group, the blue for parent's head movements, and the green for child target object perception group. 7 and 8 are child's holding and parent's holding actions, respectively. The ellipses show the 1.5σ(standard deviation of the group) equidistance trace from the group centers.

Child Perception and Head Movements

From Figure 2, we observe that the distances from child's perception to two head movement groups are changing over time. In successful learning, the child's perception (green) is closer to the parent's head movement (blue) before the naming events. A closer distance in MDS indicates that two groups were closely tied as there were more information exchanges between the two groups. This suggests that the child's perception was strongly influenced by the parent's head movements right before naming in successful learning.

One possible explanation is that in successful learning, parents either followed the child's attention or successfully attracted the child's attention. This close coupling between the parent's head movements and the child's perception served a precursor for successful learning.

However, during the naming moments, the child's perception cluster (green) moved toward the child's head movement cluster (red), and then finally moved away from both head movement clusters in Figure 2 with approximately equal distances to both, suggesting that the child's head movements were closely coupled with the child's own perception when the child heard the target object name in successful cases.
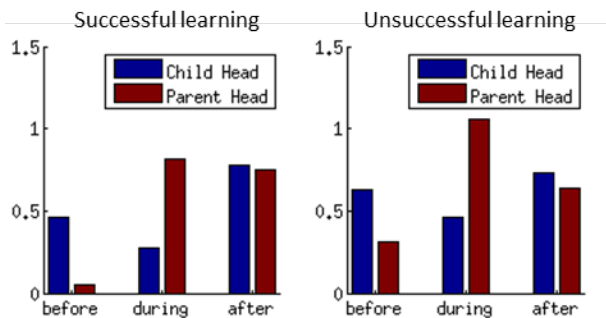

Figure 3. Closest distances between child's perception and heads movements: (Left) successful learning, (Right) unsuccessful learning.

These distance patterns are quantified and summarized in Figure 3. We calculated the distances between the groups based on the closest distance between two member variables of two groups, as a close distance between any variable pair from the two groups indicates a link between two groups through those two variables. As shown in Figure 3, the same trend described in successful cases also appeared in unsuccessful cases but the pattern is much weaker than what happened in successful moments.

Head Movements and Holding Actions

Figure 4 (top) shows the distances between the child's head movements and both the child's and the parent's holding actions. In successful cases (left), both the child's and the parent's hand actions are directly linked to the child's head movements only during naming moments but not before or after naming, suggesting a coupling of the child's head movements and manual actions from both participants at the naming moments. A similar pattern also appeared in unsuccessful learning, suggesting that the pattern is characteristic for naming moments, being successful or not. Moreover, one noticeable difference between successful and unsuccessful learning is that the parent's hand actions consistently exchanged more information with the child's head movements during and after naming in unsuccessful cases, compared with those information exchanges between the child's own holding actions and his own head movements. This pattern suggests that the child's head movements, as an indicator of the child's sustained

attention, were more influenced by the parent's (but not the child's own) manual actions, which turned out to have negative effects on learning.

The bottom two plots in Figure 4 showed information exchanges between the parent's head movements and hand actions from both participants. During naming moments, there were more information flows between the parent's head movements and manual actions in successful cases than those in unsuccessful cases. A direct coupling between the parent's head movements with manual activities can be viewed as a metric of the parent's engagement in interacting with the child and in teaching object names.
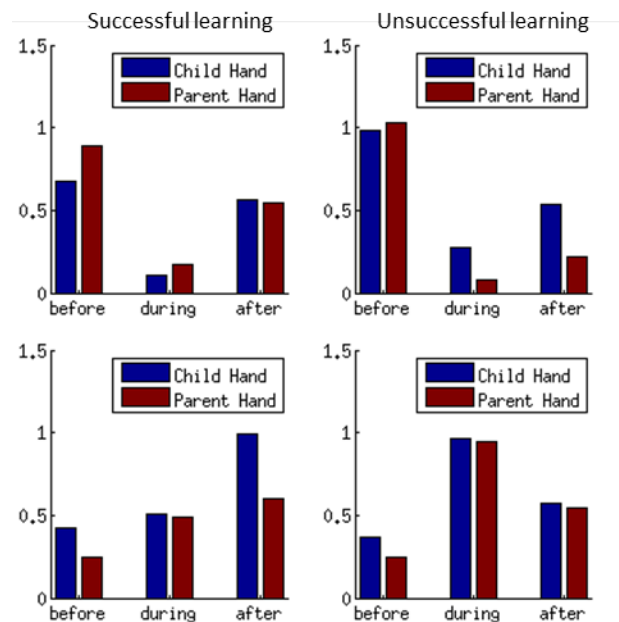

Figure 4. Closest distances between head and hand movements: (Top Left) between child's head movement and hand movements in successful learning, (Top Right) between child's head movement and hand movements in unsuccessful learning, (Bottom Left) between parent's head movement and hand movements in successful learning, and (Bottom Right) between parent's head movement and hand movements in unsuccessful learning.

Child Perception and Holding Actions

Yu, Smith, Shen, Pereira, and Smith (2009) reported that hand movements, especially the child's hand movements, are dominant factors for child's perception. However, it is not clear how their relations are measured in terms of transfer entropy. As shown in Figure 5, the overall pattern between holding actions and child perception is that before naming they are close, and then child perception moves away from holding actions during the naming moments and then returns back to hand actions after the naming moments. Moreover, this trend is shown in both successful and unsuccessful learning cases. This is a rather unexpected result as our previous studies (Yu, Smith & Pereira, 2008) showed that the child's holding actions during the naming moments can facilitate learning. Taken together with the results illustrated in Figures 3 and 4, one plausible explanation is that during those naming moments, the

child's and the parent's manual actions had direct effects (as measured by transfer entropy) on both the child's and the parent's head movements which consequentially influenced the child's perception. Thus, both the child's and the parent's head movements played a role during the naming moments as a link between the child's perception and manual actions. This additional involvement of head movements may play a critical role in naming and therefore learning. In contrast, right before and right after naming moments, head movements were not involved and therefore there were more direct information exchanges between the child's perception and manual actions.
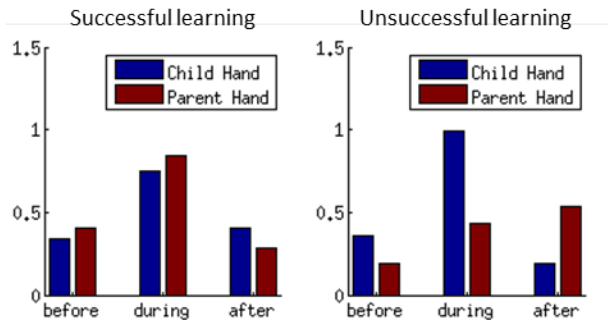

Figure5. Closest distances between hands and child's perception: (Left) unsuccessful learning, (Right) successful learning.

Discussions

The present study applied normalized transfer entropy measures to child-parent interaction data to better understand sensorimotor dynamics in such multimodal interaction. We analyzed information flows between various variables and reported how information is flowing in the child-parent interaction, especially around the naming events. Indeed, those information flows between sensorimotor variables informatively reveal a set of underlying perceptual and motor patterns which shed light on our understanding on real-time sensorimotor dynamics, within one participant and between two social partners, that lead to smooth interaction and therefore successful learning. However, we also note that, in this present approach, there are some issues that should be handled carefully. For example, there is a tendency to over-interpret directional information-theoretic measures, such as transfer entropy, because information flow itself is not causality (though it seems like causality). Experimental studies are needed to determine the causal mechanisms through which variables exchange information.

A general goal of the present study has been exploring a new venue to introduce information theoretic measures to social and behavioral studies. This approach has already been widely used in many other scientific fields and it allows the quantitative statistical analysis of many disparate systems in a mathematically rigorous way which has special merits to understand and ground high-level social interaction at the sensorimotor level. In addition, it provides a framework to study and compare seemingly different systems using the same quantitative concepts. The present

study represents our first steps to combine multimodal fine-grained behaviors and information-theoretic measures to better understand coordinated behaviors.

References

Choi, H., & Choi, S. (2007). Robust kernel Isomap. *Pattern Recognition*, 40(3), 853–862.

DeRuyter van Steveninck, R., Lewen, G. D., Strong, S. P., &Koberle, R. (1997). Reproducibility and variability in neural spike trains. *Science*, 275:1805-1807.

Gourevitch, B., & Eggermont, J. J. (2007). Evaluating information transfer between auditory cortical neurons. *Journal of Neurophysiology*, 97, 2533–2543.

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency based visual attention for rapid scene analysis. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 20(11), 1254–1259.

Lin, J., Keogh, E., Lonardi, S., & Chiu, B. (2003).A symbolic representation of time series, with implications for streaming algorithms. *In Proc. 8th ACM SIGMOD workshop on research issues in data mining and knowledge discovery*. San Diego, CA.

Marko, H. (1973). The bidirectional communication theory-a generalization of information theory. *IEEE Trans. on Communications*, COM-21(12), 1345–1351.

Massey, J. L. (1990). Causality, feedback and directed information. *In Proc. int'l. symp. on information theory and its applications*. Waikiki, Hawaii.

Schreiber, T. (2000). Measuring information transfer. *Physical Review Letters*, 85(2), 461–464.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27, 379-423.

Smith, L. B., Yu, C., & Pereira, A. F. (2010). Not your mothers view: the dynamics of toddler visual experience. *Developmental Science*, 1–9.

Sporns, O. and Lungarella, M. (2006). Evolving coordinated behavior by maximizing information structure. *In Rocha, L. et al. eds. Artificial Life X. Cambridge, MA: MIT Press*.

Tomasello, M. and Farrar, M. (1986). Joint attention and early language. *Child Develop.*,1454–1463.

Yu, C., Smith, L. B. & Pereira, A. F. (2008). Grounding Word Learning in Multimodal Sensorimotor Interaction. *In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.) Proc. of the 30th annual conference of the cognitive science society*. Austin, TX: Cognitive Science Society. 1017-1022.

Yu, C., Smith, L. B., Shen, H., Pereira, A. F., & Smith, T. G. (2009). Active information selection: Visual attention through the hands. *IEEE Trans. on Autonomous Mental Development,* 2, 141–151.