This article was downloaded by:[Pereira, Alfredo F.]
On: 24 May 2008
Access Details: [subscription number 793295812]
Publisher: Taylor & Francis
Informa Ltd Registered in England and Wales Registered Number: 1072954
Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK

Connection Science

Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713411269

Social coordination in toddler's word learning: interacting systems of perception and action

Alfredo F. Pereira [a]; Linda B. Smith [a]; Chen Yu [a]
[a] Department of Psychological and Brain Sciences and Cognitive Science Program, Indiana University, Bloomington, IN, USA

Online Publication Date: 01 June 2008

To cite this Article: Pereira, Alfredo F., Smith, Linda B. and Yu, Chen (2008) 'Social coordination in toddler's word learning: interacting systems of perception and action', Connection Science, 20:2, 73 — 89

To link to this article: DOI: 10.1080/09540090802091891
URL: http://dx.doi.org/10.1080/09540090802091891

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.informaworld.com/terms-and-conditions-of-access.pdf

This article maybe used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Connection Science
Vol. 20, Nos. 2–3, June–September 2008, 73–89

Taylor & Francis
Taylor & Francis Group

Downloaded By: [Pereira, Alfredo F.] At: 14:29 24 May 2008

Social coordination in toddler's word learning: interacting systems of perception and action

Alfredo F. Pereira*, Linda B. Smith and Chen Yu

Department of Psychological and Brain Sciences and Cognitive Science Program, Indiana University, Bloomington, IN, USA

We measured turn-taking in terms of hand and head movements and asked if the global rhythm of the participants' body activity relates to word learning. Six dyads composed of parents and toddlers ($M = 18$ months) interacted in a tabletop task wearing motion-tracking sensors on their hands and head. Parents were instructed to teach the labels of 10 novel objects and the child was later tested on a name-comprehension task. Using dynamic time warping, we compared the motion data of all body-part pairs, within and between partners. For every dyad, we also computed an overall measure of the quality of the interaction, that takes into consideration the state of interaction when the parent uttered an object label and the overall smoothness of the turn-taking. The overall interaction quality measure was correlated with the total number of words learned.

In particular, head movements were inversely related to other partner's hand movements, and the degree of bodily coupling of parent and toddler predicted the words that children learned during the interaction. The implications of joint body dynamics to understanding joint coordination of activity in a social interaction, its scaffolding effect on the child's learning and its use in the development of artificial systems are discussed.

Keywords: embodied social cognition; parent–child dyad; motion-tracking analysis; word learning

1. Introduction

A fundamental problem in understanding and building intelligent systems is the coordination of joint activity by multiple agents. Much work in this domain, both with respect to artificial and natural intelligent systems, concerns how one agent 'reads' the mind of the other. Most approaches assume that this is done by building internal models about the intentional states of the social partner and by making inferences about the goals and plans of the other (Breazeal and Scassellati 1998; Butler and Brooks 2000; Baldwin and Moses 2001; Dautenhahn and Werry 2004). Some researchers have hypothesised that these inferences are driven by perceivable body actions, presumably tied to the internal states of the actor (e.g. Smith and Breazeal 2007). In the work presented in this paper, we step back from the hypothesised mental states of the participants and their inferences about each other's internal states and attempt to directly study the dynamics of the bodily interactions themselves. The work is specifically concerned with social turn-taking – essential to conversation and collaboration.

*Corresponding author. Email: afpereir@indiana.edu

ISSN 0954-0091 print/ISSN 1360-0494 online
© 2008 Taylor & Francis
DOI: 10.1080/09540090802091891
http://www.informaworld.com

Downloaded By: [Pereira, Alfredo F.] At: 14:29 24 May 2008

In human adults, the shifting of turns in a conversation is exquisitely fine-tuned and so flawless that participants seem to shift roles instantaneously. A smooth interaction is also one with overlaps and with no gaps (e.g. Coates 1994). Moreover, the perceived quality of the interaction is related to the seamless quality of shifting turns (see Wilson and Wilson 2006 for a review). We seek to understand the nature of the bodily dynamics that determine these interactions. Our approach has three components. First, we study turn-taking in parent interactions with their toddlers. Adult turn-taking is so skilled and rapid that the underlying components are hard to discern (Wilson and Wilson 2006). By examining turn-taking when one partner is less skilled, we may be able to disentangle the component threads that are so tightly woven in mature interactions as well as understand the developmental origins of turn-taking. Secondly, we seek to relate the dynamics of turn-taking to the effectiveness of the social interaction. We operationally define effectiveness in terms of how much a toddler learns about the names of objects heard during the interaction. The assumption is that an effective interaction between a parent and toddler is one that promotes learning. Finally, we measure the dynamics in terms of the global rhythm of head and hand movements. This is based on the past research with adults suggesting that whole body rhythms orchestrate the timing of role shifts in adult conversation (Shockley, Santana and Fowler 2003) and our own recent studies suggesting that such bodily rhythms are also evident in parent–child interactions (Yu et al. under review).

1.1. *A body rhythm*

There is considerable research both with children and with artificial agents on the meaning of such specific actions as pointing or eye gaze direction (Brooks and Meltzoff 2002). However, quality social interactions, which yield successful information exchange, require more than cues that signal meanings. They also require precise timing. Moreover, this timing must be managed locally by the participants, turn by turn. In mature partners, simultaneous talk and simultaneous silences occur rarely and briefly (see Wilson and Wilson 2006, for an extensive review). The question, then, is how does one participant know when to listen or watch and when to speak or act? The precision in the timing of adult role shifting suggests an underlying rhythm, or oscillation, that patterns opportunities for shifting. Oscillatory properties of internal cognitive processes (Buzaki and Draguhn 2004) or perhaps even breathing patterns (McFarland 2001) may contribute, but more extensive research suggests that participants use a consortium of cues including hand and body movements (Dittman and Llewelly 1968; Harrigan 1985; Walker and Triboli 1982; Shockley, Santana, and Fowler 2003; Wilson and Wilson 2006). Moreover, these results suggest that the oscillatory rhythm itself may be emergent in the coupled body actions of the participants and dependent on context, emotion and culture (see, for example, Beattie 1979; Street 1984; Coates 1994; Jungers, Palmer and Speer 2002; Murata 1994; Shockley et al. 2003). In a recent review, Richardson, Dale and Kirkham (2006) concluded that events such as naming or pointing may be dynamically embedded in this larger bodily rhythm (Kita 2003; Bangerter 2004). Accordingly, in the present study, we measure turn-taking in terms of the shift in amounts of hand and head movements between social partners. The goal here is not to determine the potential 'meaning' of any individual movement, but the rhythm of the activity among participants and its relation to learning. Does a better rhythm lead to better learning? An affirmative answer would suggest that the timing of attention and perhaps the binding of cognitive contents are determined by the social rhythm itself.

1.2. *Scaffolding by the mature partner*

Research from parent–infant interactions suggests that the more mature partner may play a critical role in orchestrating or controlling this rhythm. Studies of early face-to-face play of parents and 2–3 month-old infants indicate that parents impose a rhythm of doing and observing on young

Downloaded By: [Pereira, Alfredo F.] At: 14:29 24 May 2008

infants (e.g. Cohn and Tronick 1998; Rogoff 1990; Schaffer 1996; Trevarthen 1988). For example, when the baby coos, the mother watches, then when the baby goes still, the mother coos. The result is a joint pattern of activity and orienting (Beebe, Stern, and Jaffe 1979). This pattern has generated considerable interest because (1) it has the same general structure as conversations and (2) because the mature partner in the interaction seems to treat the interaction as a structured exchange of information, perhaps in this way, teaching the rhythm of information exchange. These observations also suggest how a turn-taking rhythm could be controlled by very simple local processes – of stopping activity when the other is acting (or when there is something interesting to watch) and initiating activity when the other is not moving (see Capella 1981 for a review). Such a turn-taking rhythm would emerge – though perhaps with a few flaws – even if only one partner followed these simple rules (see Yu, Smith and Pereira under review).

Such turn-taking – even if controlled by one partner – might give the surface appearance of information exchange, but is it enough to yield the successful exchange of information? There is at present no data on this possibility for parent–child social interactions. Research with adults suggests that coordinated bodily actions among participants leads to both better bonding (Oullier, de Guzman, Jantzen, Lagarde, and Kelso 2007; Newtson, Hairfield, Bloomingdale, and Cutino 1987) and better memory for the event (Lagarde and Kelso 2006). Other research with humans (Large and Jones 1999; Gogate, Bahrick, and Watson 2000) as well as with artificial agents (Fitzpatrick and Arsenio 2004) suggests more generally that multimodal rhythms can support the integration of information and learning. Thus, the functional value of scaffolded turn-taking may include the rhythmic organisation of attention to support learning. Accordingly, we ask if the dynamics of parent–child turn-taking predict object name learning by the child.

1.3. *Word learning in artificial agents*

Learning how to associate spoken words with their referents poses a challenge to artificial intelligence systems, in parallel to the difficult question of explaining this capability in humans (Quine 1960). This is because of the noise and uncertainty inherent in real world learning with its many words, many co-occurring referents, and shifts in attention. In the area of embodied artificial intelligence systems, studies using real sensors have achieved some success (Steels and Kaplan 2001; Steels and Vogt 1997; Cangelosi and Parisi 2002; Roy and Pentland 2002; Yu, Ballard, and Aslin 2005). However, those systems can only learn a certain kind of words (object names or colour names), and it is non-trivial to extend to more general cases and open cases. One important difference between human learning and machine learning systems is that people learn language mostly in social contexts through everyday teacher–learner interactions. In contrast, most artificial systems acquire knowledge based on logic inferences and statistical computation. Nevertheless, some systems do consider the role of social cues in language learning. For example, Yu and Ballard (2005) used a speaker's eye movements to infer the referential intent. Steels and Vogt (1997) and also Steels and Kaplan (2001) developed robotic systems showing how symbols can emerge from a communication consensus gradually built up in a social interaction. There are also studies of artificial systems that demonstrate how the dynamic coupling of multiple agents may yield rhythms of coordination and cooperation (Marocco, Cangelosi, and Nolfi 2003; McMillen, Rybski, and Veloso 2005; Di Paolo, Rohde, and Iizuka (in press)). However, there has not been systematic consideration of the role of social rhythms such as those that underlie turn-taking in supporting learning.

Our approach to this problem is to collect fine-grained sensory data from parent–child interactions and analyse and model the social interaction patterns that create the turn-taking rhythm. This knowledge should be useful in building artificial systems that engage in the most effective way with both humans and other artificial systems.

Downloaded By: [Pereira, Alfredo F.] At: 14:29 24 May 2008

2. Measuring joint body dynamics in toddler–parent dyads

Toddlers have conversations and learn words in contexts that involve more than just words and labelled objects (Huttenlocher 1974; Goldin-Meadow, Seligman and Gelman 1976; Gogate et al. 2000). The natural learning context consists of sequences of actions, often on objects, and shifts in attention among objects. The standard approach to understanding the structure of these interactions in developmental psychology is to code by hand experimenter-defined behaviours such as looks, gestures, touches of objects and so on (Ahktar 1996; Dromi 1999; Iverson and Capirci 1999). There are several problems with this approach. One is these human coders may only notice and categorise behaviours consistent with their own already developed notions of intention and goal and not note other behaviours that may play a critical role in regulating the interaction. A second problem is that these kinds of discrete descriptions about individual acts cannot provide fine-grained dynamic information about the flow of activity itself. If we are going to understand how coupled rhythms of activity may support learning and information exchange, we need a finer-grained measure of the temporal properties of the interaction.

To achieve this goal, we used motion-tracking equipment to collect the dynamic data on the real-time movements of parents and toddlers as they were engaged in a naturalistic task of conversation during toy play. By placing motion sensors on the heads and hands of the participants, we collected continuous high-resolution data about their movements. To the best of our knowledge, this is the first study to use this method to understand how body movements relate to learning in parent–child dyads.

2.1. *Toy play and measuring object name learning in a parent–child interaction*

Six parents and their children (three male, three female) participated. The children ranged in age from 17 to 19 months ($M = 18.2$ months). One additional child was recruited, but did not tolerate the measuring devices and did not contribute data. The period between 17 and 20 months is one of considerable vocabulary development in young children, with children's productive vocabulary increasing from as few as 10 words to as many as 250 (Fenson et al. 1994). Children of the age of those in this study are thus, in their everyday life, very much engaged in learning new words.

The experimental task was one common to the everyday lives of children, one in which children and parents take turns in jointly acting on, attending to and naming objects. This is a common context in which children learn names for things. The toys (Figure 1) used in this experiment were everyday things, but things for which very young children (according to normative measure of typical early vocabularies, Fenson et al. 1994) are unlikely to know the names: comb, butterfly, bulldozer, rolling pin, cube, cricket, pliers, donkey, racket, and machine. The child and parent

Figure 1. Stimuli used in the toy play and word learning experimental study: cricket, butterfly, comb, cube, machine, rolling pin, donkey, pliers, racket, and machine. The black line close to each object is one inch in length.

Downloaded By: [Pereira, Alfredo F.] At: 14:29 24 May 2008

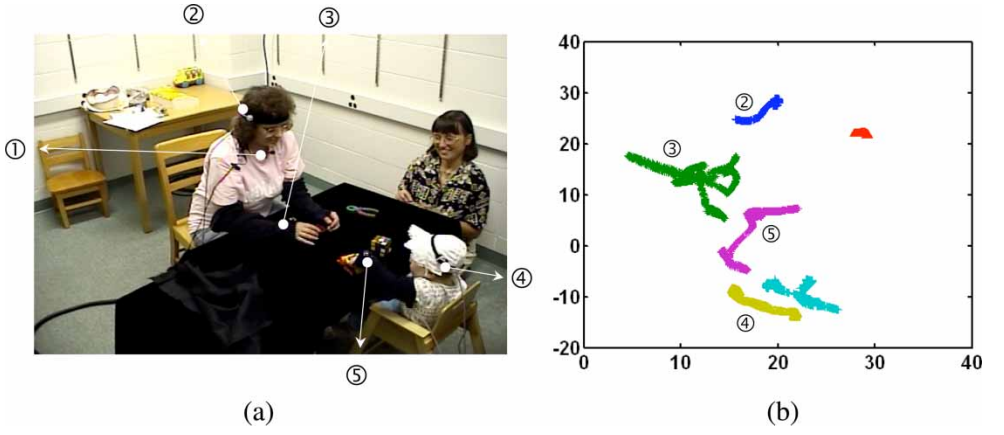

Figure 2. a) View of the experimental set-up. The parent wore a headset microphone (1), one motion-tracking sensor placed on a headband (2) and one on each hand (3). In addition, the child also wore a motion-tracking sensor placed on a headband (4) and one on each hand (5); b) Example of movement data from the two participants, projected on the table plane.

sat opposite to each other at a small table (61 × 91 × 64 cm) and the parent was instructed to interact naturally with the child, engaging their attention with the toy. The parent was told the experimenter-selected names of each object (those listed above) and was asked to use that name when referring to the object.

We placed motion sensors (described below) on the hands and heads of each participant. The motion sensors for the head were embedded in a sports headband. The hand sensors were sewn into fingerless gloves (see the experimental set-up in Figure 2). The procedure to place the sensors used two experimenters: upon entering the room, the parent and child sat down and one experimenter introduced the child to an engaging pop-up toy; while the child was fully engaged with this toy, a second experimenter placed the headband on the child. The experimenter then helped the parent place hers. The parent placed the fingerless glove on the child. Putting the sensors on the two participants took 5 min. The parent wore a headset with a microphone from which we recorded the parent's voice during play.

The interaction between parent and child lasted between 7 and 12 min and was free-flowing in form. The session began with three toys on the table. Periodically, when interest in a toy waned, the experimenter offered new toys to the parent such that over the course of the session all children were exposed to 10 toys and their names.

After this period of interaction, the experimenter took the parent's place across from the child at the table and tested the child's comprehension of the object name for each of the 10 objects. This was done by placing two objects out of reach of the child about 30 inches apart, one to the left of the child and one to the right. Then looking directly into the child's eyes, the experimenter said the name of one of the objects and asked for it (e.g. 'I want the truck! The truck! Get me the truck!'). For this portion of the experiment, a camera was focused on the child's eyes. The direction of eye gaze – looking at the named object when named – was scored as indicating comprehension. Eye movement data was coded (with the sound-off) by a scorer naïve to the experiment. In addition, a human coder also coded the mother's recorded speech. The coder listened to the recording and time-stamped (onset and offset) any utterances of the 10 object labels included in the study.

2.2. *Tracking head and hand motion*

To measure the activity of each partner's head and hands, we used an electromagnetic motion-tracking solution, the Liberty system from Polhemus. This tracker uses passive electromagnetic

Downloaded By: [Pereira, Alfredo F.] At: 14:29 24 May 2008

sensors and a source that emits a magnetic field. The source was placed under the table. The sensors consist of electromagnetic coils in a plastic casing, assembled as small cubes measuring $22.9 \times 27.9 \times 15.2$ mm and weighing 23 g. A wire connects each sensor to the base and multiple sensors can be acquired simultaneously with high-sampling rates and precision. When tracking, the system provides six DOF data, 3D coordinates (x, y, z), and 3D orientation (heading, pitch, and roll) relative to the source position. A total of six sensors were used: one for the head and each hand of the two participants. An example of raw data collected in this study, namely the 3D position of head and hands for both parent and child, is given in Figure 3.

2.3. *Data analysis methodology*

Our main goal was to investigate the nature and quality of the global rhythm of the interaction. Thus, the data analysis was focused on reducing the multidimensional time series into a compressed form so that each partner's activity level could be directly compared. By reducing the dimensionality of the raw data, we tested if overall activity, irrespective of detailed position or orientation, can be used as an indicator of the quality of interaction and of joint coordination. This was done by comparing the amount of body-part motion, after dimensionality reduction, across partners within a dyad. In addition, we used this basic idea to correlate the interaction's overall quality, particularly those aspects indicative of turn-taking between parent and child.

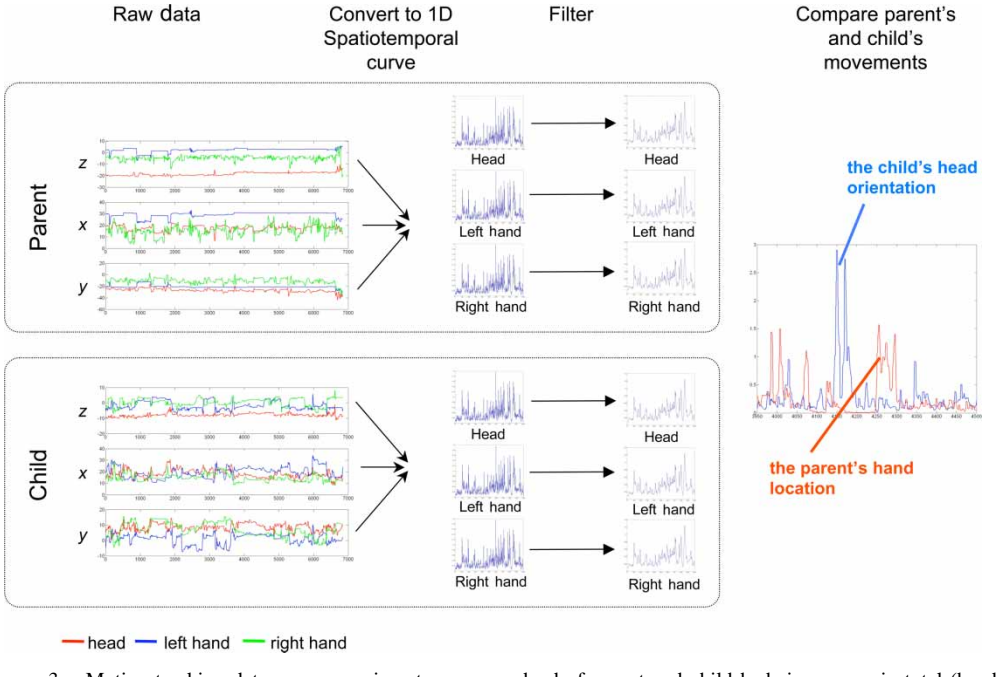

Figure 3. Motion-tracking data pre-processing steps: every dyad of parent and child had six sensors in total (head, left and right hands), and each sensor outputs a six-dimensional vector (three-dimensional position and orientation), the figure shows on the left an example of raw position data from one dyad; the position and orientation data is reduced to a one-dimensional spatiotemporal, separately for each, and the resulting signal is filtered; each dyad generates 12 unidimensional signals; finally we align all pairs of body-part activity using dynamic time warping and calculate the cost of alignment. On the right is an example is related to turn: the child's head orientation is compared to the caregiver's hand position.

Downloaded By: [Pereira, Alfredo F.] At: 14:29 24 May 2008

In summary, our approach requires the following steps: data reduction and filtering (for simplicity of the analysis); a method to compare the level of body-part activity between partners; a measure of the overall quality of the interaction. We describe each of these components before discussing the experiment's results.

Data reduction and filtering

The first step in measuring joint body dynamics in parent–child dyads was reducing the three-dimensional position data or the three-dimensional orientation data of each partner's head and left and right hands into a one-dimensional measure. We used the multidimensional spatiotemporal curvature technique (Rao, Yilmaz, and Mubarak 2002). This transformation preserves information on variations in speed, direction, and acceleration. This curve measures the overall activity that can be achieved by any combination of axial motion. The transformation can also be applied to the position data (x, y, z) or the orientation data (h, p, r) of any sensor, thus reducing each sensor to a unidimensional signal. The transformation applied was as follows:

$$k = \frac{\sqrt{A^2 + B^2 + C^2 + D^2 + E^2 + F^2}}{((x')^2 + (y')^2 + (z')^2 + (t')^2)^{3/2}}$$

$$A = \begin{vmatrix} y' & t' \\ y'' & t'' \end{vmatrix}, \quad B = \begin{vmatrix} t' & x' \\ t'' & x'' \end{vmatrix}, \quad C = \begin{vmatrix} x' & y' \\ x'' & y'' \end{vmatrix},$$

$$D = \begin{vmatrix} z' & t' \\ z'' & t'' \end{vmatrix}, \quad E = \begin{vmatrix} z' & x' \\ z'' & x'' \end{vmatrix}, \quad F = \begin{vmatrix} z' & y' \\ z'' & y'' \end{vmatrix},$$

where x, y, z are the three-dimensional position data or alternatively the h, p, r orientation data. The variable t is the datapoint's time-stamp and the primes denote, for example, in the case of x, $x'(t) = x(t) - x(t-1)$ and $x''(t) = x'(t) - x'(t-1)$. Each one-dimensional signal was smoothed and high-frequency components removed using a standard Kalman filter (Haykin 2001). Given the large number of datapoints (sampling rate was 100 Hz), a single set of parameters was estimated for the Kalman filter and used for all signals. After these two steps, each parent–child dyad generated 12 time series corresponding to parent's head position and orientation (two), parent's left and right hand position and orientation (four), children's head position and orientation (two), and children's left and right hand position and orientation (four). This process is summarised in Figure 3.

Comparing activity of body parts across partners

With the dimensionality reduced data, we can more easily compare the bodily activity across the two dyad members. If the social partners are jointly modulating their behaviour, there should be a correlation between a partner's head or hands activity and the other partner's head or hands motion. More specifically, a negative correlation would mean that the increase of activity in one partner leads to a deceleration of the other partner's activity and vice versa, which is an indicator of turn-taking. A positive correlation would mean that the variation of activity in one partner leads to similar variation of activity in the other; that is they are synchronised.

However, to adequately compare body activity across partners, we must consider that each partner's movements naturally overlap and the initiation of action by one may be followed by a change in activity of the other partner, but only after a variable time delay. Given that the social interactions we are studying include a child, this issue is increased since time delays are of a different magnitude between partners (adults are faster to adjust motion, for example). Computing

Downloaded By: [Pereira, Alfredo F.] At: 14:29 24 May 2008

a correlation directly in time or inside a sliding window of a fixed size thus may not capture the interaction correctly. More generally, this corresponds to the problem of computing a similarity measure between two time series. A pair of time series may share similar features that however do not align in the time dimension (or addition in the Y dimension). To compensate for this, we used dynamic time warping (DTW) to compare two time series (Rabiner and Juang 1993, Berndt and Clifford 1994; Koegh and Ratanamahatana 2004). This technique aligns two time series by warping non-linearly in time, one series into the other; searching for a warping path that minimises a cost function. The resulting alignment cost can be used as a dissimilarity measure. DTW has been extensively used in speech recognition as a solution for time normalisation (stemming from the fact that utterances of a token by the same speaker are never exactly the same in speed) and speech recognition, using the cost of the DTW alignment to search for the most similar token (Rabiner and Juang 1993). Current applications of DTW include data mining, gesture recognition, robotics, speech processing, and medicine (Koegh and Ratanamahatana 2004).

Briefly, DTW's algorithm searches for a warping path between time series Q and time series C in the following manner: the distances between any datapoints q_i and c_j (usually the Euclidean distance is used) are stored in a $q_m x c_n$ matrix, where m and n are the sizes of Q and C; a path through this matrix corresponds to one particular alignment of the two series; a path's cost is measured by the sum of the point-to-point distances; DTW searches for a path that minimises this cost, subject to a set of constraints (e.g. boundary, continuity, monotonicity); current versions of the algorithm can search for the best warping path quite efficiently. For more details, see, for example, Koegh and Ratanamahatana 2004.

The DTW cost of alignment can be used to compare all pairwise combinations of body-part activity (in the present study, head and left and right hands, compared within and between dyad members). A low cost of alignment means that the warping path resides mainly close to the diagonal; the minimum cost warping path is then one where each point is aligned to another close to it in time. This happens when two series are reasonably synchronised or at least positively correlated inside small time windows. The higher the cost of this alignment, the further from the diagonal is the minimum cost warping path, that is each point is aligned to another that is distant in time. In the data from this study, this happens when the moments of activity in one partner correspond to moments of inactivity in the other partner and vice versa so that the best alignment is always distant in time. This last scenario corresponds to the partners engaging in more turn-taking. An example of these two scenarios is in Figure 4.

The DTW cost of alignment produces a measure without easily interpretable units, so we calculated a baseline to allow for relative comparisons across dyads. We created random pairings of children and parents; afterwards data from the child participant and parent participant for that random pairing was aligned using dynamic time warping. If a specific pairwise comparison inside a true parent–child dyad is statistically higher than the averaged equal comparison in all the random pairings (e.g. head orientation with other partner's hand movements), it would mean that the activity in the real dyad tends to be temporally in a different direction to the activity on the other. Lower than baseline means that the two signals are more similar than in randomly aligned series.

Measuring the overall quality of an interaction

The DTW alignment measures when the activity overlaps for a large portion of the interaction or each partner tends to stop moving when the other partner is acting. Such comparisons of bodily activity across partners can be used to calculate a summarised overall index of the entire interaction that captures the quality of the social exchange in terms of smoothness of turn-taking. To this measure, we added the effectiveness of the interaction for promoting learning, which in turn was

Downloaded By: [Pereira, Alfredo F.] At: 14:29 24 May 2008

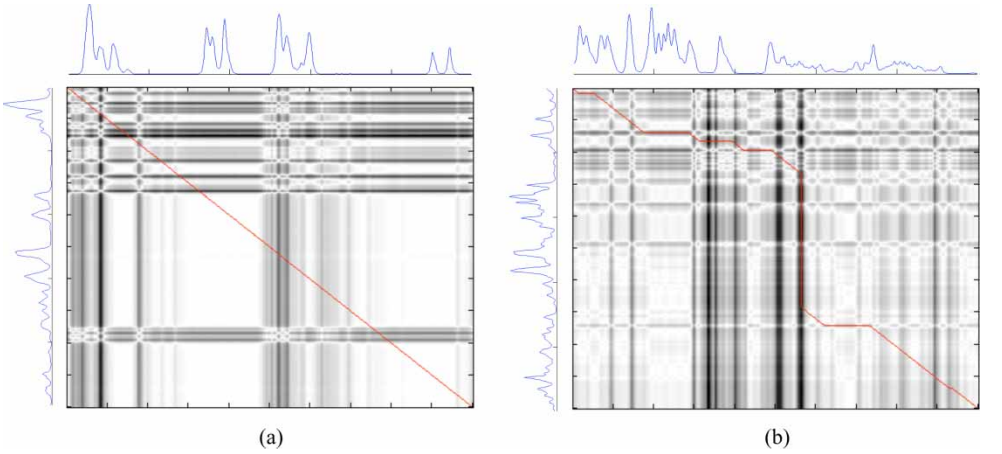

Figure 4. Two contrasting scenarios of motion-tracking data: in (a) activity in one time series tends to be matched by activity in the second; in (b) activity in one usually corresponds to inactivity in the paired time series. The matrix shown in each panel depicts all point-to-point costs (darker corresponds to smaller cost) and the minimum cost warping path, as found using dynamic time warping, is overlaid in red.

measured as the number of words comprehended by the child in the testing phase. We evaluated the turn-taking metric (a proxy of the smoothness in the cycling of activity between partners) with the word testing measure by zooming in on periods of the interaction where the parent uttered the object labels we previously selected. This is critical because an interaction could appear overall to have a good joint coordination of action but could, nonetheless, be in a disorganised state (e.g. both participants moving) during the key event of the parent's utterance of an object name. In natural interactions, key events are almost never in perfect cross-modal synchrony; sometimes the child is playing while the parent labels the object, other times the child observes the parent when the word is said, but may also be coactive during the naming event. A one-word utterance can also overlap several interaction states, as motor movements damp down or increase during the utterance. Our measure takes all this into consideration by first marking every data point, inside the utterance of an object label, with an interaction state. We categorised the participant's joint-action states into: (1) both still: child and parent were not moving; (2) caregiver lead: the caregiver was moving while the child was not; (3) child lead: the child's body was moving while the caregiver was not moving; and (4) concurrent: both were moving. This last state would seem to create difficulties to the success of information exchange in the interaction. For a data point to count as moving, the only requirement was that any body part (head, left or right hand) needed to be moving. To be categorised as not moving, no body part could be moving.

The final aggregate measure calculates a weighted average of the time spent on each possible interaction state, giving more weight to a state of non-coactivity such as (2) or (3) and weighted this average by the overall turn-taking measure. We combined states (2) and (3) into one state called turn-taking and only considered the relations between activity in a partner's head and the other partner's hands as these are the strongest indicators of turn-taking. The next equation shows the calculation:

$$\text{Overall Quality}_{\text{Dyad } A} = \overline{DTW}_{\text{Dyad } A} \sum_{i=1}^{j} \sum_{t=1}^{k} \text{state\_dyad}(i,t)$$

$$\overline{\text{DTW}}_{\text{Dyad } A} = \frac{\text{DTW}(\text{Child's Head, Parent's Hands}) + \text{DTW}(\text{Parent's Head, Child's Hands})}{2}$$

Downloaded By: [Pereira, Alfredo F.] At: 14:29 24 May 2008

$$\text{state\_dyad}(i,t) = \begin{cases} 1.0, & \text{if dyad was in a turn} - \text{taking state at time t} \\ 0.5, & \text{if dyad was in both still state at time t} \\ 0.0, & \text{if dyad was in concurrent state at time t} \end{cases}$$

$$i = 1 \ldots j^{th} \text{ utterance of an object label by parent}$$

$$t = 1 \ldots k^{th} \text{ timestamp in utterance i}$$

$$\text{DTW} = \text{dynamic time warping cost of alignment}$$

3. Observed patterns of body dynamics

The data from each dyad (six sensors plus the mother's speech) was analysed using the steps described previously. We found that (1) these measures are sensitive to turn-taking; (2) parents and children were actively turn-taking; and (3) the quality of the turn-taking interaction is correlated with the number of objects' labels that children learned.

3.1. *Comparison of body activity between partners*

We aligned, using DTW, all possible pairwise comparisons of the three measured body parts (head, left and right hands) within and between the participants. As described in the previous section, to assess the degree of coordination within a dyad, we compared this measure to a baseline composed of random parent–child pairings, testing the average DTW alignment cost of the sample with the average DTW alignment of the random pairings sample for each specific comparison. When the comparison included hand movements, both the left and right hands were included in the comparison. This was done by first computing the DTW alignment for each hand with another body party separately and then including the results from both hands when comparing against baseline. For example, when comparing head orientation movements to hand location movements, the sample of six true dyads generated 12 DTW alignment costs (one head orientation compared with the left and right hands) and the six random dyads generated another 12 comparisons; when testing against baseline each set of 12 points was one group.

The results of the DTW comparisons were as follows: (1) child's head orientation and the parent's hand movements had a higher cost (indicating turn-taking) than baseline for DTW alignment, $t(22) = 4.29$, $P < .001$, one-tailed t-test, and calculated over the independent DTW alignments of the parent's left and right hands to the child's head; (2) the parent's head orientation and the children's hand movements had a higher cost than baseline for DTW alignment (again, indicative of turn-taking), $t(22) = 3.12$, $P < 0.005$, one-tailed t-test; (3) both the child's and the caregiver's own head orientation compared with their own hand movements had a lower cost than baseline for DTW alignment (indicating synchronisation of the activity of different body parts within an individual), $t_{\text{child}}(22) = 2.75$, $P < .005$, $t_{\text{parent}}(22) = 2.86$, $P < 0.005$, both one-tailed t-tests. These results indicate that (1) children tend to decrease head motion when the parent is moving the hands, (2) parents tend to decrease head motion when the child is moving the hands, and (3) the participants own head and hands movement were coordinated. All other pairwise comparisons were not statistically significant ($P > 0.05$).

Most relevant to the present goals are the first two results. These indicate that both child and adult jointly coordinated their own body activity to the activity of the social partner. Head movements were inversely related to other partner's hand movements, i.e. the head was still while the partner's hands were moving. This is a strong indicator of turn-taking. Importantly, then, turn-taking is evident in the global activity pattern of body movement itself.

Downloaded By: [Pereira, Alfredo F.] At: 14:29 24 May 2008

3.2. *Quality of the social interaction and its effect on word learning*

On an average, parent–child interactions in this context of toy play and word learning exhibit turn-taking. However, across dyads, interactions varied considerably in the degree of the coordination smoothness, i.e. in the joint timing or cycling of body activity and inactivity. If it is the case that the quality of the social interaction, in terms of the consistency in cycling body activity, is relevant to children's word learning, then a measure of turn-taking smoothness should be related to how much was learned during that social interaction.

Accordingly, we calculated for each dyad an overall quality of the interaction, using the metric described earlier in Section 2.3, and compared it with the number of words the children comprehended when tested by the experimenter. Figure 5 shows the results of this comparison. We found that dyads with the highest interaction quality resulted in a higher total number of objects correctly selected by the child in the name-comprehension task. That is, the smoother the bodily activity modulation between the social partners, the more effective was the interaction for the child. Again, it is noteworthy that this result was detected with a measure of the whole interaction, calculated only from motion-tracking data of body movements, which suggests a causal role of joint coordination of body dynamics in orchestrating successful information exchanges.

The interaction quality measure weighs the turn-taking by the state of interaction during a target word utterance. This does not tell us if either the parent or the child was still when the word was uttered and how much of the interaction during a word utterance is a non-concurrent state. It could be that dyads with the most effective interactions were also in specific activity–inactivity states when the parent uttered the object label. If so, how strong is the bias towards certain interaction modes? The distribution of interaction states for the time periods where an object label was uttered is shown in Figure 6 for each individual dyad. The name-comprehension result is also marked. With the present sample size, it is not possible to derive strong generalisations, but the differences between the dyads are suggestive of preferential modes of organising the interaction, as it relates to word learning. The dyadic interactions yielding the highest child performance in the name-comprehension task had the least uniform distributions. They also showed a high probability for the interaction to be in the child-lead or the parent-lead state during labelling; that is, the parent said the word when the child was not moving (perhaps indicating that the child stops to listen), or the parent uttered the word when not moving herself (perhaps providing a verbal description of the child's actions). That is, within the dyadic interactions yielding the most word learning by

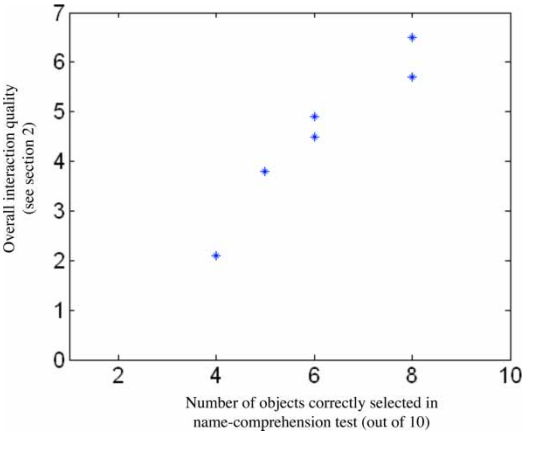

Figure 5. Overall interaction quality measure (as described in Section 2) and total number of objects correct in the name-comprehension test for each dyad.

Downloaded By: [Pereira, Alfredo F.] At: 14:29 24 May 2008

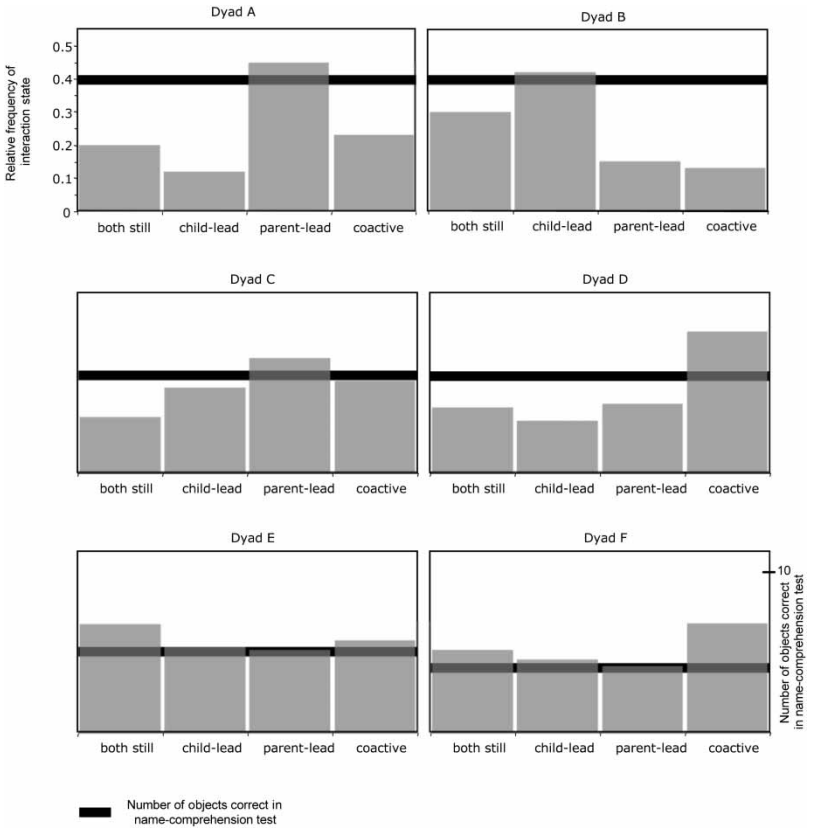

Figure 6. Distribution of the frequency of the parent–child interaction state, calculated over the interaction time corresponding to all the utterances of an object label, for each dyad. The total number of objects correctly selected in the name-comprehension test (out of 10) is also shown below each distribution plot as a thick line. The dyads are ordered by decreasing value of the name-comprehension test. For each axis, all six panels use the same scale.

the child, naming events by parents did not occur when both participants were moving nor when both were still.

The individual patterns also suggest that multiple strategies can lead to a successful joint coordination and information exchange, perhaps a reflection of the idiosyncrasies of one's own body activity and capability of reading body activity in others. This may be particularly so in parent–child combinations. Nonetheless, some patterns appear less likely to lead to learning. The dyads yielding the poorest performance in the name-comprehension task had very uniform distributions with little evidence of preferences for the interaction mode in which only one participant was moving. These interactions were clearly the least smooth, where the cycling of activity between parent and child was not consistent and most importantly for the child's learning, the interaction was not consistent when key events for learning occurred, that is, an object label was introduced.

4. General discussion

A few minutes of toy play between parent and toddler yields a plethora of activity. In all its complexity, the exchange entails many perception-action possibilities: head turns, hand motion and hand gestures, eye gaze shifts, language, properties of the speech signal, affect, objects seen, objects held, and so on. The present results show that all those individual acts create a global

Downloaded By: [Pereira, Alfredo F.] At: 14:29 24 May 2008

rhythm of activity–inactivity that cycles between the two participants, with momentary overlaps and synchronies of perception and action. The overall structure of the rhythm observed here, though coarser, approximates the finely tuned social exchanges of adults. Moreover, the results show that this global rhythm – tracked in the present study only through head and hands motion – matters to learning. Thus, the main result is this: there is a cycling in the global rhythms of parent–child body activity and inactivity that organise toddler's learning from that interaction. How and in what way might this global cycling of activity support object name learning? We offer two, not mutually exclusive, hypotheses.

Convergent and temporal cues to joint attention

Considerable research in learning object names, both with children and with artificial systems, has demonstrated the importance of the child's attention to the speaker's intended referent. That attention to the object needs to occur when the speaker names the referent. Picking out the right referent at the right moment is not a trivial problem in cluttered real-world learning tasks, with many shifts in attention and intended referents as discourse proceeds. Consequentially, considerable research has been concerned with how children use individual acts by the speaker, such as pointing or, more frequently, direction of eye gaze, to find the intended referent. Most of this research with children and with artificial agents have used discrete and long-lasting acts of looking and pointing to a referent. This is necessary, in part, because the geometrical calculations to actually discern the referent from such acts are not simple and certainly not simple for young children (Triesch, Teuscher, Deak, and Carlson 2006). This fact raises the serious problem as to how children actually use these acts (or how artificial devices can be made to use these acts) to discern the referent in the highly dynamic context of continuous discourse. However, if these acts – and naming – are systematically embedded in a global rhythm of activity and inactivity, then sensitivity to global activity itself (rather than individual acts) or the rhythm itself may support joint attention and perhaps also attention to specific kinds of body movements. An important future question, then, is how joint attention may be dynamically related to the turn-taking rhythm seen in global activity.

Following or leading the learner

The naming events most predictive of word learning occurred: (1) when the child was acting and the parent was still and (2) when the child was still and the parent was acting. The first context may emerge when the parent's attention (and naming) follows the child's attentional lead. Studies of individual differences in how parents teach children as well as experimental studies have led some to conclude that the most efficient teaching episode is one in which the teacher names the object to which the child is already attending (see, for example, Tomasello and Todd 1983; Tomasello, Mannle and Kruger 1986; Tomasello and Farrar 1986). Indeed, in one longitudinal study, toddlers learned more words when taught using the follow the child's lead approach than in conditions in which the child's attention was recruited to the referent before teaching (Tomasello and Farrar 1986). Here, the child does not have to figure out the referent because the mature partner waits until the child's attention is directed to one thing and then names that thing of interest. The second context – naming an object when the child is still and watching – could be viewed as an instance of the mature partner trying to lead attention. However, these naming episodes may also be effective if they emerge from parents waiting to act until when the child's own activity is low and naming after the child directs attention to the parent's action. A global cycling of activity and inactivity may be coupled with a give and take in attention that yields optimal and participatory learning. Certainly, naming events in which both participants are acting – and neither is still and

Downloaded By: [Pereira, Alfredo F.] At: 14:29 24 May 2008

watching the attentional state of the other – seem suboptimal. Naming when no one is acting – and thus when attention may be less engaged – may also be suboptimal. A key question for future work is whether both cycles (child active/parent inactive and child inactive/parent active) are equally effective for learning and whether the rhythm of global activity plays an additional role in orchestrating the timing of these teaching opportunities.

4.1. *Scaffolding an immature social partner by shaping his internal dynamics*

Adult human social interactions show an impressive range of properties that are matched and regulated. Child and adult social exchanges show part of this structure, but they have an important qualitative difference: they are between a mature and a considerably more immature partner. The immature partner in these interactions has a smaller set of motor skills, lower attention span, less cognitive control, more difficulty in selectively attending, and engaging and disengaging attention. The interaction is thus asymmetric, in striking contrast to a social interaction between two mature partners that share a considerable set of cognitive abilities and cultural background, which can be used to reach a common ground.

It seems likely that in the present context, the mature partner orchestrated the turn-taking rhythm. Just as in the turn-taking of mother–infant face-to-face play, it may be the mature partner who stops what she is doing if the child starts to act and who increases her activity on objects when the child's own activity begins to wane. In another related study, Yu, Smith and Pereira (under review) examined the structure of transitions among four possible states in the interactions of toddlers and parents. The states were identical to the definitions we used here: (1) both still: neither is moving; (2) caregiver lead: the caregiver is moving while the child is not moving; (3) child lead: the child's body is moving while the caregiver is not moving; and (4) concurrent: both are moving. Transitions to states 2 and 3 were primarily controlled by the parent's activity. When both were still, the parent began to increase activity; when both were concurrently acting, the parent stopped acting. Importantly, by controlling the dynamics of the interaction in this way, the mature partner may shape or entrain the internal dynamics of the immature one. Developmental advances in information exchange and collaborative problem-solving thus may have their origins in the tuning of attentional and activity-based rhythms (see Tani, Nishimoto, Namikawa, and Ito 2008 for evidence suggesting that social partners can 'force' a developing system into more advanced intrinsic dynamics).

4.2. *Social learning in human–machine interaction*

A deeper understanding of human learning is directly relevant to building artificial intelligent systems that learn from, teach, and work with humans. Decades of research in artificial intelligence suggest that flexible adaptive systems cannot be fully pre-programmed. Instead, we need to build systems with some preliminary constraints that can create and exploit a rich and variable learning environment. Considerable advances have been made in biologically inspired forms of artificial intelligence (Brooks, Breazel, Irie, Kemp, and Marjanori 1998; Asada, McDorman, Ishiguro, and Kuniyoshi 2001; Steels and Kaplan 2001; Breazeal and Scassellati 2002; Weng, Zhang and Yu 2003; Yu and Ballard 2005).

Toddlers are fast word learners and they do so from interactions with people and objects in a cluttered world. Could we build a computational system that accomplishes the same learning task? If so, what attributes of a young child are crucial for the machine to emulate? We believe that studies in human learning provide useful hints in various aspects to answer those questions. First, human studies suggest what kinds of technical problems need to be tackled. For the same task, mapping language to the real world, the machine needs to deal with similar problems that the

Downloaded By: [Pereira, Alfredo F.] At: 14:29 24 May 2008

existing intelligent systems, biological or not, face. For example, one key problem in human language learning is reference uncertainty (Quine 1960); given a natural learning situation consisting of a sequence of spoken words uttered by a teacher with multiple co-occurring objects and events in the extra-linguistic context the present work shows how joint body activity in a social interaction may orchestrate attention and thus help to resolve this ambiguity and promote language learning. Moreover, the current study focused on a social context common to young children, but quite different from current machine learning (ML) approaches. Many ML approaches first collect data with (or without) teaching labels from users and the environment and then rely on implementing efficient mathematical algorithms and applying them onto the pre-collected data to induce language knowledge. The methodology largely assumes that a learner (e.g. a machine) passively receives information from a language teacher (e.g. a human supervisor) in a one-way flow. In contrast, a young child is situated in social contexts and learns language through his own actions with objects and the caregiver. Language teachers dynamically adjust their behaviour based on their understanding of the learner's state. Thus, teachers provide 'on-demand' information in real-time learning. Meanwhile, the learner also plays an important role in learning-oriented interactions by actively generating actions to interact with the physical environment and to shape the teachers' responses and acquire just-in-need data for his learning.

Thus, current machine learning studies focus on one aspect of learning – what kind of learning device can perform effective computations on the pre-collected data; they ignore an equally important aspect of the learning – the learning environment that a learner is situated in. The present results hint that two dynamically coupled systems, who cycle through periods of activity and inactivity, of doing and watching, may create an even more powerful learning device than the child (or machine) alone. If the rhythm in an interaction can gate attention and learning, then a key mechanism for learning is outside the individual learner and resides in the dynamic coupling of the learner to a social world.

Acknowledgements

This research was supported by grants from NSF (BCS0544995) and NIH (R21 EY017843) awarded to the second and third author. The first author was also partially supported by a Fulbright fellowship and a PhD scholarship from the Gulbenkian Foundation. The authors wish to thank Charlotte Wozniak for help in data collection.

References

Akhtar, N. (1996), "The Role of Discourse Novelty in Early Word Learning," *Child Development*, 67, 2.

Asada, M., MacDorman, K., Ishiguro, H., and Kuniyoshi, Y. (2001), "Cognitive Developmental Robotics as a New Paradigm for the Design of Humanoid Robots," *Robotics and Autonomous Systems*, 37, 185–193.

Baldwin, D., and Moses, L. (2001), "Links between Social Understanding and Early Word Learning: Challenges to Current Accounts," *Social Development*, 10, 309–329.

Bangerter, A. (2004), "Using Pointing and Describing to Achieve Joint Focus of Attention in Dialogue," *Psychological Science*, 15, 415–419.

Beattie, G.W. (1979), "Contextual Constraints on the Floor – Apportionment Function of Speaker-Gaze in Dyadic Conversations," *British Journal of Social and Clinical Psychology*, 18, 391–392.

Beebe, B., Stern, D., and Jaffe, J. (1979), "The Kinesic Rhythm of Mother-Infant Interactions," in *Of Speech and Time,* eds. A. Steigman and S. Feldstein, Hillsdale, N.J.: Erlbaum, pp. 23–24.

Berndt, D., and Clifford, J. (1994), *Using Dynamic Time Warping to Find Patterns in Time Series. AAAI-94 Workshop on Knowledge Discovery in Databases (KDD-94)*, Seattle, Washington.

Breazeal, C., and Scassellati, B. (2002), "Robots that Imitate Humans," *Trends in Cognitive Sciences*, 6, 481–487.

Breazeal, C., and Scassellati, B. (1998), "Infant-like Social Interactions Between a Robot and a Human Caretaker," *Adaptive Behavior*, 1, 49–74.

Brooks, R.A., Breazeal, C., Irie, R., Kemp, C.C., and Marjanovi, M. (1998), "Alternative Essences of Intelligence," in *AAAI '98/IAAI '98: Proceedings of the Fifteenth National/Tenth Conference On Artificial Intelligence/Innovative Applications Of Artificial Intelligence*. Menlo Park, CA, USA: American Association for Artificial Intelligence, pp. 961–968.

Downloaded By: [Pereira, Alfredo F.] At: 14:29 24 May 2008

Brooks, R., and Meltzoff, A.N. (2002), "The Importance of Eyes: How Infants Interpret Adult Looking Behavior," *Developmental Psychology* 38, 958–956.
Butler, S., Caron, A., and Brooks, R. (2000), "Infant Understanding of the Referential Nature of Looking," *Journal of Cognition and Development*, 1, 359–377.
Buzsaki, G., and Draguhn, A. (2004), "Neuronal Oscillations in Cortical Networks," *Science*, 304, 1926–1929.
Cangelosi, A., and Parisi, D. (eds.) (2002), *Simulating the Evolution of Language,* London: Springer-Verlag, pp. xii+355.
Coates, J. (1994), "No Gap, Lots of Overlap: Turn-taking Patterns in the Talk of Women Friends," in *Researching Language and Literacy In Social Context,* eds. D. Graddol, J. Maybin and B. Stierer, Avon, UK: Mulitilingual Matters Ltd, pp. 177–192.
Cohn, J.F., and Tronick, E.Z. (1998), "Mother-Infant Face to Face Interaction: Influence is Bi-Directional and Unrelated to Periodic Cycles in Either Partner's Behaviour," *Developmental Psychology*, 24, 386–392.
Dautenhahn, K. and Werry, I. (2004), "Towards Interactive Robots," *Pragmatics and Cognition,* 12, 1–18.
Di Paolo, E.A., Rohde, M., and Iizuka, H. "Sensitivity to Social Contingency or Stability of Interaction?," *Modelling the Dynamics of Perceptual Crossing New Ideas in Psychology Special issue on Dynamics and Psychology*, (in press, ww.sciencedirect.com/science/article/B6VD4-4PJ6BMX-1/1/).
Dittmann, A.T., and Llewellyn, L.G. (1968), "Relationship between Vocalizations and Head Nods as Listener Responses," *Journal of Personality and Social Psychology*, 9, 79–84.
Dromi, E. (1999), "Early Lexical Development" in *The Development of Language*, M.D. Barret (ed.), Hove, UK: Psychology Press.
Fenson, L., Dale, P.S., Reznick, J.S., Bates, E., Thal D.J., and Pethick, S.J. (1994), "Variability in Early Communicative Development," *Monographs of the Society for Research in Child Development*, 59 (5, serial 242).
Fitzpatrick, P., and Arsenio, A. (2004), "Feel the Beat: Using Cross-Modal Rhythm to Integrate Perception of Objects, Others and Self," in *Fourth International Workshop on Epigenetic Robotics*, Genoa, eds. L. Berthouze, H. Kozima, C. G. Prince, G. Sandini, G. Stojanov, G. Metta and C. Balkenius, pp. 59–66.
Gogate, L.J., Bahrick, L.E., and Watson, J.D. (2000), "A Study of Multimodal Motherese: The Role of Temporal Synchrony between Verbal Labels and Gestures," *Child Development*, 71, 878.
Goldin-Meadow, S., Seligman, M.E.P., Gelman, R. (1976), "Language in the two-year-old," *Cognition*, 4, 189–202.
Harrigan, J.A. (1985), "Listener's Body Movements and Speaking Turns," *Communication Research*, 12, 233–250.
Haykin, S. (2001), *Kalman Filtering and Neural Networks*, NJ: Wiley-Interscience.
Huttenlocher, J. (1974), "The Origins of Language Comprehension," in *Theories in Cognitive Psychology,* ed. R. Solso, Hillsdale, NJ: Erlbaum, pp. 331–368.
Iverson, J.M., Capirci, O., Longobardi, E., and Cristina, M. (1999), "Gesturing in Mother-Child Interactions," *Cognitive Development* 14, 57–75.
Jungers, M.K., Palmer, C., and Speer, S.R. (2002), "Time After Time: The Coordinating Influence of Tempo in Music and Speech," *Cognitive Processing*, 1, 21–35.
Keogh, E. and Ratanamahatana, C.A. (2004), "Exact Indexing of Dynamic Time Warping," *Knowledge and Information Systems*, 7, 358–386.
Kita, S. (ed.) (2003), "*Pointing: Where Language, Culture, and Cognition Meet*," Mahwah, NJ: Erlbaum.
Lagarde, J., and Kelso, J. (2006), "Binding of Movement, Sound and Touch: Multimodal Coordination Dynamics," *Experimental Brain Research*, 173, 673.
Large, E.W., and Jones, M.R. (1999), "The Dynamics of Attending: How People Track Time-Varying Events," *Psychological Review*, 106, 119–159.
Marocco, D., Cangelosi, A., and Nolfi, S. (2003), "The Emergence of Communication in Evolutionary Robots", *Philosophical transactions: Mathematical, Physical and Engineering Sciences*, 361, 2397–2421.
McFarland, D.H. (2001), "Respiratory Markers of Conversational Interaction," *Journal Speech Language and Hearing Research*, 44, 128–143.
McMillen, C., Rybski, P., and Veloso, M. (2005), "Levels of Multi-robot Coordination for Dynamic Environments," in *Multi-Robot Systems: From Swarms to Intelligent Automata,* Vol. 3, eds. L.E. Parker. F.E. Schneider and A. Schulz, Netherlands: Springer, pp. 53–64.
Murata, K. (1994), "Intrusive or Co-operative? A Cross-Cultural Study of Interruption," *Journal of Pragmatics*, 21, 385–400.
Newtson, D., Hairfield, J., Bloomingdale, J., and Cutino, S. (1987), "The Structure of Action and Interaction," *Social Cognition*, 5, 191–237.
Oullier, O., de Guzman, G., Jantzen, K., Lagarde, J., and Kelso, J.A.S. (2007), "Social Coordination Dynamics: Measuring Human Bonding," *Social Neuroscience*, 1.
Quine, W. (1960), *Word and Object*, Cambridge, MA: MIT Press.
Rabiner, L., and Juang, B. (1993), *Fundamentals of Speech Recognition*, Upper Saddle River, NJ: Prentice Hall.
Rao, C., Yilmaz, A., and Mubarak, S. (2002), "View-Invariant Representation And Recognition of Actions," *International Journal of Computer Vision*, 50, 203–226.
Richardson, D., Dale, R., and Kirkham, N. (2007), "The Art of Conversation Is Coordination: Common Ground and the Coupling of Eye Movements During Dialogue," *Psychological Science*, 18, 407–413.
Rogoff, B. (1990), *Apprenticeship in Thinking: Cognitive Development in Social Context*, Oxford: Oxford University Press.
Roy, D., and Pentland, A. (2002), "Learning Words from Sights and Sounds: A Computational Model," *Cognitive Science*, 26, 113–146.
Schaffer, H.R. (1996), *Social Development*, Oxford: Blackwell.

Downloaded By: [Pereira, Alfredo F.] At: 14:29 24 May 2008

Shockley, K. Santana, M.V., and Fowler, C.A. (2003), "Mutual Interpersonal Postural Constraints are Involved in Cooperative Conversation," *Journal of Experimental Psychology: Human Perception and Performance*, 29, 326–332.

Smith, L.B., and Breazeal, C. (2007), "The Dynamic Lift of Developmental Process," *Developmental Science*, 10, 61–68.

Steels, L., and Kaplan, F. (2001), "Aibo's First Words; the Social Learning of Language and Meaning," *Evolution of Communication*, 4, 3–32.

Steels, L., and Vogt, P. (1997), "Grounding Adaptive Language Game in Robotic Agents," in *Proceedings of the Fourth European Conference on Artificial Life,* eds. C. Husbands and I. Harvey, London: MIT Press.

Street, R.L. (1984), "Speech Convergence and Speech Evaluation in Fact-Finding Interviews," *Human Communication Research*, 11, 139–169.

Tani, J., Nishimoto, R., Namikawa, J., and Ito, M. (2008), "Co-developmental Learning between Human and Humanoid Robot Using a Dynamic Neural Network Model," *IEEE Trans. on Systems, Man, and Cybernetics Part B: Cybernetics*, Vol. 38, January.

Tomasello, M., and Farrar, M.J. (1986), "Joint Attention and Early Language," *Child Development,* 57, 1454–1463.

Tomasello, M., Mannle, S., and Kruger, A. (1986), "The Linguistic Environment of One to Two Year Old Twins," *Developmental Psychology*, 22, 169–176.

Tomasello, M., and Todd, J. (1983), "Joint Attention and Lexical Acquisition Style," *First Language*, 4, 197–211.

Trevarthen, C. (1998), "Infants Trying to Talk," in *Children's Creative Communication,* ed. R Söderbergh, Lund: Lund University Press.

Triesch, J., Teuscher, C., Deak, G., and Carlson, E. (2006), "Gaze Following: Why (Not) Learn it?," *Developmental Science*, 9, 125–147.

Walker, M.B., and Triboli, C. (1982), "Smooth Transitions in Conversational Interactions," *The Journal of Social Psychology*, 117, 305–306.

Weng, J., Zhang, Y., and Yu, C. (2003). "Developing Early Senses about The World: 'Object Permanence' and Visuoauditory Real-Time Learning," in *Proceedings in International Joint Conference on Neural Networks,* Portland, pp. 2710–2715.

Wilson, M., and Wilson, T.P. (2006), "An Oscillator Model of the Timing of Turn-taking," *Psychonomic Bulletin and Review*, 12, 957–968.

Yu, C., Ballard, D.H., and Aslin, R.N. (2005), "The Role of Embodied Intention In Early Lexical Acquisition", *Cognitive Science*, 29, 961–1005.

Yu, C., Smith, L.B., and Pereira, A.F. "Body Prosody: the Rhythm in Parent-Child Social Interactions", unsubmitted publication, copy with author, *Infancy*.