Interactional Structure Applied to the Identification and Generation of Visual Interactive Behavior: Robots That (Usually) Follow the Rules

Bernard Ogden[1], Kerstin Dautenhahn[1], and Penny Stribling[2]

[1] Adaptive Systems Research Group, Department of Computer Science, University of Hertfordshire, Hatfield, Herts, UK
{bernard|kerstin}@aurora-project.com,
Project home page: http://www.aurora-project.com
Department home page: http://homepages.feis.herts.ac.uk/nehaniv/ASRG.html
[2] School of Psychology and Counselling, Whitelands College, University of Surrey Roehampton, West Hill, London, UK
P.Stribling@btinternet.com

Abstract. This chapter outlines the application of interactional structures observed by various researchers to the development of artificial interactive agents. The original work from which these structures are drawn has been carried out by researchers in a range of fields including anthropology, sociology and social psychology: the 'local approach' described in this paper draws particularly on conversation analysis. We briefly discuss the application of heuristics derived from this work to the development of an interaction tracking system and, in more detail, discuss the use of this work in the development of an architecture for generating action for an interactive agent.

1 Introduction

This paper discusses the relevance of work from the human sciences (i.e. fields such as anthropology, sociology and social psychology) to the development of systems or agents that are concerned with observing or with engaging in interaction. Our primary interest is in the latter of these two types of artificial agent: however, we believe that the work is also relevant to identification of interaction and we discuss this briefly in section 2. Section 3 considers the question of generation of appropriate behaviors for an interactive agent, drawing heavily on concepts from the field of conversation analysis (CA). Finally, we briefly consider the issue of assessing and measuring interaction in section 4 before concluding.

Before proceeding, it is necessary to clarify what we mean by interaction. We see interaction as a reciprocal activity in which the actions of each agent influence the actions of the other agents engaged in the same interaction, resulting in a mutually constructed pattern of complimentary behavior.[1] Key in this

[1] For those readers who are familiar with the framework of autopoiesis our definition of interaction might be reminiscent of the notion of 'structural coupling' which de-

definition is the co-constructed nature of the interaction: this aspect of interaction is emphasized in Clark's notion of joint action [3] and in di Paolo's work with interacting virtual agents [4]. At the present time we are most interested in visual interaction, by which we mean an interaction in which action is detected primarily in the visual channel. In other words, we are interested in interaction in which the actions of participants are detectable by sight, as in the case of movement, gesture and the like.[2]

We also wish to note that we are discussing systems that deal with the output from an underlying machine vision system: in principle, the ideas discussed here should be compatible with any machine vision system capable of identifying actions of interest and labeling its identifications with a probability of correctness. Finally, we should mention that we are primarily interested in interactive *robots*: thus we often refer to robots rather than artificial agents more generally in the following discussion. We nonetheless believe that this discussion is relevant to any artificial agent that is intended to interact with humans, including virtual agents e.g. embodied conversational agents [5, 6].

2 Identification of Global Structures

In this section we consider structures of interaction that operate at a global level[3] i.e. structures that are applied to relatively long sequences of interaction rather than operating at a more local level.[4]

An example of a global structure would be a greeting as described by [7]: here we have the idea of a greeting as composed of a number of phases, from distant salutation through approach to a final close salutation. Each phase in this structure has a set of typical associated actions. We briefly discuss here means of identifying such a sequence if it is observed as part of a vision sequence — the interested reader is referred to [2] for a more detailed discussion of the same ideas. We note that the ideas in this section draw particularly on the work of Kendon [8, 9] which, in turn, reflects the familiar AI/cognitive science notion of scripts [10]. We also note that [11] describes similar work to Kendon in this respect.

For our purposes we consider a global structure as being composed of a number of phases, as noted above, each with a set of associated actions. Each action has a meaning for each phase in which it occurs: thus if an action occurs

---

scribes the relationship between a living organism and its environment [1]. Indeed, we explore these similarities in more detail in [2].

[2] Note that we are not claiming that visual resources are involved in the *production* of such actions, which will usually take the form of movement and gesture. This is merely a term for describing the channel through which the actions are detected.

[3] Note that the ideas in this section are mainly *not* drawn from conversation analysis, which generally sees conversation as locally managed.

[4] In [2] we discuss large and small *units* of interaction rather than long and short *sequences* of interaction. We now prefer the term 'sequences' as it implies a less rigid unit and thus is a better description of the case in human interaction.

in more than one phase it can have a different meaning in each phase (e.g. a 'wave hello' vs. a 'wave goodbye'). We view the problem of identifying an action sequence of a given interaction as one of assigning probabilities of being an instance of a given interaction type to each observed potential action sequence. We assume that the basic machine vision system itself is able to assign some kind of base probability to each action it observes and we take this probability as our starting point in selecting an action sequence. We can then consider heuristics based on the structure of human interaction to weight these probabilities.[5]

Continuous Phase Progression Heuristic. It seems that a sequence that proceeds through each and every phase from beginning to end should be weighted more heavily than one that skips some phases of the preferred sequence or ends without reaching the end of the preferred sequence, although both of these events happen in the course of normal interaction [7, 12]. Some 'deviations' from the preferred sequence may be more common than others and this should be reflected in weights assigned according to this heuristic. We note that variations in the preferred sequence in normal interaction will in themselves be actions in the interaction and thus of interest — however, in the present case we are interested in classifying an observed sequence of behaviour as belonging or not to one member of a set of classes of types of interaction (e.g. greeting, argument, making a purchase). We are therefore working at a very coarse granularity and not attempting to determine the full implications of everything that we see.

Globally Improbable Phase Transition Heuristic. A further heuristic is simple weight of numbers. If we observe three actions in phase x of an interaction, then an action in phase $x + 1$, then a further four actions in phase x, we should probably conclude that the transition to phase $x + 1$ never happened and that the observation of the action was spurious.

Adjacency Pair Heuristic. In cases where a given action creates the expectation of a given response (e.g. Kendon [7] notes that a head nod is often responded to with another head nod) then interpretations of the data that show this pair should be weighted more strongly than alternative interpretations.

Contiguous Action Heuristic. In a turn-taking interaction it seems reasonable to expect minimal overlapping and leaving of gaps between consecutive actions most of the time [13, 12, 14].[6] In such interactions we can assign greater

[5] We note that these heuristics are not necessarily to be used in a real-time system. Some are simply inappropriate for this purpose: for example, the boundary signaling heuristic will involve reinterpretation of assumptions regarding prior actions and hence potentially will require reanalysis of the whole sequence of actions-so-far several times over in the course of a single sequence of interaction.

[6] Hutchby and Wooffitt [14] also points out that in certain types of interaction there may be more gap and overlap. It is also worth noting that gaps can be acts in an

probability to interpretations that minimize overlapping and gaps between the actions of interactants.[7]

Boundary Signaling Heuristic. The actions of the first phase of an interaction define its beginning: similarly, the actions of the last phase of an interaction define its end. This allows us to reduce the set of action sequences of interest to those that begin with a first phase action of some interaction and to weight action sequences that terminate with an action from the end phase of a matching interaction more heavily than those that do not.

3 Local Structures

We now move from structures that organize longer sequences of interaction to those that organize shorter sequences. These local structures[8] often work at the level of action and response[9] and so seem especially suited to incorporation into systems that generate interactive action as they can provide a means for selecting an action based only on the prior action and, possibly, on the likely nature of the next action. This minimizes the past memory or future planning required by the agent, allowing considerably computational savings. We examine structures observed by conversation analysts in their studies of human interaction and consider the application of similar concepts to a robotic agent. To provide some context we consider the case where these structures are applied to the agent in the 'dancing with strangers' experiment [15], but note that these interactional structures could be applied in architectures other than the very simple one considered here. More generally, CA has already been applied in the development of new interfaces by researchers in human-computer interaction [16].

3.1 The Dancing Robot

Dautenhahn [15] describes an experiment in which a robot coordinates its movements to a human's, modifying its movement behaviour in response to reinforce-

interaction in themselves [13, 12] but, as noted above, we are dealing with a coarse grained classification rather than a detailed analysis at this stage.

[7] Note that rules of conversation, a special case of auditory interaction, will not necessarily correspond to the case of visual interaction. In this case, the visual channel seems much more open to overlapping actions than the auditory channel - an individual who speaks while someone else is speaking creates interference in the channel, but it is certainly possible to move and gesture simultaneously with another individual without necessarily causing interference. Thus the value of this heuristic is uncertain at this time.

[8] We note that what we describe as 'local' and 'global' structures are both seen as locally managed in the CA view. Here we are using the term 'local' to describe an approach that deals with short sequences of interaction (as opposed to a 'global' approach involving longer sequences of interaction), rather than in its CA sense.

[9] Note, though, that the response is itself an action in the interaction i.e. a response is itself an action requiring a response in turn.

ment from the human's hand movements which are classified into six categories: left, right, up, down, clockwise and anti-clockwise. Our discussion here is confined to the 'autonomous-select' condition of the experiment: in this case the robot cycles through a few basic behaviors with the human's hand movements selecting one or a few of these behaviors by reinforcement.[10] The experiment employs an association matrix relating inputs and outputs. A weight in the matrix is activated when the two agents perform the matching behaviors i.e. the weight that exists for the pair c_1/r_1 where c_1 is the human's movement input and r_1 the robot's movement output is activated when the human performs movement c_1 and the robot performs movement r_1. A weight is increased if it is activated in two consecutive time steps and decreased when it is not activated. This condition is called temporal coordination between the movements of human and robot. In this way, using a simple reinforcement mechanism, the robot will 'learn' to perform particular movements in response to particular human movements and a simple interaction can develop between robot and human. It is easy to imagine extensions of this experiment in which the movements of interest are not necessarily the six categories of hand movement but instead are any set of arbitrarily defined actions.

3.2 The Interactively Active Robot

Interaction is often viewed as actively constructed by the interactants: this is also the case in our own definition in section 1. In the basic 'dancing' case as described above, however, the interaction is entirely led by the human interactant: the robot's only contributions are simple responses. We could get closer to 'real' interaction by having the robot more actively constructing the interaction. To achieve this the robot must have goals in the interaction. Ordinarily a human might approach an interaction with various goals in mind. The robot's goals will obviously be much simpler than those in the human case: specifically, they can be the production of some specific sequence of actions on the part of both interactants. In order to be able to achieve its goals the robot needs to understand the likely effects of its own actions on the human: to this end it needs to build up a set of mappings from its own behavior to the human's. This seems simply achieved by a similar system to that which constructs the mappings from the human's behavior to the robot's. To begin, we take the set R to be the set of all of the robot's possible actions and the set C to be all of the human's actions that the robot can recognize and thus respond to and we take a network in which all human actions have a weight connecting them to all robot actions and vice versa (i.e. action c_1 is connected to every action in the set R, as is c_2, c_3, c_4 etc. In turn, each action in the set R is connected to each action in the set C by a different weight). Note that weights are not symmetric i.e. the weight mapping r_1 to c_1 (w_{r_1/c_1}) is not the same as the weight mapping c_1 to r_1 (w_{c_1/r_1}). Now, instead of having all non-activated weights decay in each time step we have the weight w_{r_1/c_x} decrease in cases where action r_1 is followed

[10] The reader is referred to the original paper for more information.

instead by some other action c_y: the weight w_{r_1/c_x} would also increase, of course. In this way the weights can adapt to observed behavior. We can consider any weight above some arbitrary threshold t_w to be a valid action. Thus, we could have structures such as the following: $r_1/c_1|c_2|c_3$ where r_1 is the first part of the pair and c_1, c_2 and c_3 are the set of expected actions of the human following action r_1: of these three alternatives, one action should be produced. We then have 'action chains' composed of action/response pairs as the basic structure of our interaction where the first action of a sequence defines a set of valid responses, each of which in turn has a further set of valid responses and so on. For the robot to generate a target sequence it needs to both create a situation where the correct mappings exist for the sequence to have a high likelihood of being produced and to successfully produce the first action of the sequence. However it should not take control of the interaction to such an extent that, for example, when it has the goal sequence r_1/c_1, r_2/c_2, r_3/c_3, r_2/c_4, it produces r_1 constantly until c_1 is produced in response and so on until it eventually succeeds in forcing the interaction sequence that it was seeking (indeed, it is not necessary that it ever reaches the target sequence, but it should attempt to 'guide' the human in this direction: the final pattern of behavior will thus be an amalgam of the goals of both interactants). We can achieve a degree of control by modifying the weights mapping human action to robot response: in this way the robot acquires the ability to guide the interaction in a given direction without taking all control away from the human. Thus if we have the structure $c_1/r_1|r_2|r_3|r_4$ the weight mapping c_1 to r_2 could be multiplied by some arbitrary factor α, where $1 < \alpha$. Similarly, the other weights in this structure could be multiplied by some arbitrary factor β, where $0 \leq \beta < 1$. Exact values for α and β would have to be determined empirically, but in general the higher the value of α and the lower the value of β the more control the robot has and the less control the human has. Thus the robot's likelihood of producing a part of its goal sequence is increased without it necessarily slavishly repeating the same response on each occasion that a relevant human action is produced. We also need, of course, to consider the reverse case: the human's action produced in response to the robot's action. We obviously cannot achieve this through simple adjustment of weights, as the weights mapping robot action to human response obviously have no effect on the human's actions. One solution is to catch the case where the human coincidentally produces the correct response and attempt to reinforce this behavior, e.g. by producing some rewarding action (e.g. pretty light displays, 'happy' sounds etc.). We note that further alternative goals are possible, such as maintaining a given distance from one another (as in [4]), maintaining a given heading with respect to one another, moving to a certain area in the space within which the agents are interacting, etc. It would also be possible to introduce something similar to the homeostatic control used by the robotic head Kismet [17–19]. The precise nature of the goals is not important — what matters is that the agents be able to mutually influence one another in order to achieve these goals.

3.3 Conversation Analysis and Computational Architectures

We wish to apply conversation analytic concepts in the construction of a computational system for an interactive robot. In doing this it is necessary to consider the appropriate application of these concepts in designing artificial interactive systems. It is particularly important to understand the nature of conversation analytic 'rules'. These are not rules in the usual sense as they do not define steps that are followed by people to produce conversation but instead describe the manner in which it is produced. People can be said to 'orient to' these rules, rather than to follow them: the rules will not necessarily be followed all the time, but violations will usually be noted and inferences drawn from them.[11] Thus not following a rule is an interactive act in itself. We take the following example of the effect of violation of a CA rule (here, a perceived failure to produce an expected second part of an adjacency pair):

(Two colleagues pass in the corridor)
1 A: Hello.
2 B: ((almost inaudible)) Hi
3 (Pause: B continues walking)
4 A: ((shouts)) HEllo!

(Example from [14] p. 42). The notation conventions used in this paper are described in appendix A.)

In this example A appears not to hear B's simultaneous and quiet greeting and this violates B's expectation that the greeting will be responded to. In this case the interpretation that may be drawn from the assumed failure to complete the adjacency pair is that person A is being snubbed, or alternatively that he has not been heard. In either case, as far as A is concerned the normal expected sequence has not been followed and thus it is appropriate for A to engage in repair. Thus we can see that the rule that a greeting is followed by a greeting is violable in that A and B have interacted even though B does not realize that his greeting is being responded to, but that this violation (deliberate or otherwise) itself constitutes an action in the interaction.[12] While this example is conversational, we can consider a similar case in a 'dancing' interaction in which one robot performs an action with a given expected response which the other fails to provide (in this case, the other might fail to move at all, or else fail to change its existing movement).

We note that although the methodology of CA seems appropriate for research into visual interactive behavior and has indeed been used in this manner [20,

[11] In (more accurate) CA terms, the rules are indexical practices locally oriented to and locally produced: talk is context sensitive in that it orients to cultural notions of how a conversation should proceed but also context building in that the act of selecting particular ways of structuring talk affirms, renews or subverts the culturally given notion of structure.

[12] This example also demonstrates that a non-action (perceived or actual) can constitute an action in the interaction

21], many of CA's rules are designed to deal with conversational interaction and may not be directly applicable to visual interaction. Nonetheless, we believe that principles and observations derived from this field can be applied in our case. We are particularly interested in the concept of joint action [3], the idea of turn taking as a local management system [13] and the notion of repair [14, 3, 13]. We focus particularly on repair in this paper as human interaction is very flexible and not strictly bound by inviolable rules, which presents obvious problems for a computational controller. We see repair[13], combined with a notion of the rules that the controller uses as violable, as creating the potential for a controller that is able to use such rules without being constrained to always follow them. Before continuing with our consideration of repair, however, we would like to note advantages stemming from seeing interaction in terms of joint action and of the local management view of turn-taking. Firstly, the concept of joint action [3], (also exhibited in [4]) gives us a clearer idea of what interaction actually is and a means of determining if some form of interactive behavior has actually been achieved e.g. by contrasting the behavior of two interacting agents with the case where one agent is attempting to interact but the other is simply 'playing back' a recording of the actions of an agent in an earlier interaction [4]. Secondly, the idea of turn-taking as a local management system is also promising. The evidence that CA offers that this is how human conversational turn-taking works suggests that such a system is certainly a workable way of managing interaction (or at least that it would provide a part of a complete system for such management). This would simplify the process of managing an interaction considerably by not requiring any reference to external rules of larger scope, such as some of the heuristics introduced in section 2.[14] Local management of transitions between interactants and problems that occur reduces the need to maintain memory of preceding actions and the need to plan ahead beyond the next turn — in a global approach such planning could extend from the present point to the expected end of the interaction. These points provide obvious computational advantages.

3.4 Robotic Repair

In CA, repair is a means of correcting a misunderstanding or a mistake in an interaction, or of correcting a deviation from the normal rules of interaction. Conversation analysts generally consider four kinds of repair[15] [14, 12]. These

[13] Frohlich and Luff [22] provides an example of the use of repair in an existing computational system

[14] External context is important in CA. However, the 'context-free' aspect of certain CA concepts, such as the conversation turn-taking system described in [13] is the part that we focus on in this paper as it seems to offer the greatest computational advantages.

[15] Clark [3] describes an additional three categories of repair: preventatives, which prevent a problem from occurring; warnings, which warn of a future unavoidable problem and repairs, which refers to repair of the type considered here in which action is taken after a problem in order to resolve it. While these ideas are interesting and relevant they will not be considered here due to space limitations.

four kinds of repair are generally labelled as follows, where 'self' is the interactant whose action is being repaired and 'other' is the other interactant:

1. Self-initiated self-repair: problem detected and repaired by self
2. Other-initiated self-repair: problem pointed out by other but repaired by self
3. Self-initiated other-repair: problem pointed out by self, repaired by other (e.g. prompting for help with a memory lapse)
4. Other-initiated other-repair: problem both pointed out and repaired by other

The ability to engage in repair is essential in interaction: errors and misunderstandings are likely to arise and must be corrected if the interaction is to be successful. In applying concepts from CA in a visual interaction we might expect to face some difficulties, as CA usually deals with sequences of talk-in-interaction rather than sequences of movement. However, we find that it is possible to consider equivalent situations in a purely visual interaction. We will consider each of the above classes of repair in turn, first in cases where the human is 'self', then in cases where the robot is 'self'.

Human Error: Self-Initiated Self-Repair. Let us consider the case where the human performs action c_1 but meant to perform c_2. He might well rapidly replace c_1 with c_2, something that he could accomplish in a number of ways. He could transform c_1 to c_2 by changing action part-way through (this is likely to confuse virtually any present machine vision system); he could abort c_1 part-completed and perform c_2 instead or he could complete c_1, abort and perform c_2, immediately or after an arbitrary amount of time. In all of these cases we may expect negation behaviour of some kind, such as head-shaking or saying 'no', although such behaviour is not guaranteed to occur.[16] In each of these cases, if c_1 has been identified then the robot may already be engaging in some manner of response: this is most likely, of course, in the cases where c_1 is completed. We could build a short delay into the robot, giving it time to detect a second action following negation behaviour. If negation behaviour is detected then the robot can terminate any response that it is currently engaged in and wait for the next non-negating action. However 'negation behaviour' is a very abstract concept and could include a range of behaviors, including vocalizations, head shaking, hand waving, a pause in action and so on. It seems unlikely that a robot could be programmed to recognize visual 'negation behaviour' generally, although it could be programmed to recognize specific instances of such behaviour if particular kinds of negation behaviour are found to occur frequently.

Human Error: Other-Initiated Self-Repair. One form of repair is repetition of an action that has not produced an expected response, as in the following example from [14](p. 42):

[16] Other negation behavior is possible, as long as the robot is able to perceive and correctly interpret it. One might also think of a set of predefined and easy to recognize words/gestures, a basic 'vocabulary', possibly domain-specific, that can be used in human-robot interaction.

Child: Have to cut these Mummy. (1.3) Won't we
Mummy.
(1.5)
Child: Won't we.
Mother: Yes.

Here the child responds to the mother's failure to answer her question (a violation of adjacency pair structure) by repeating the question until getting an answer.[17] This seems a sufficiently simple behavior for the robot to be able to engage in it. Thus if the human performs an action that violates the robot's expectations (remember that inaction can be considered a form of action), repetition of the elicitor provides a means for attempting to produce the expected behavior instead.

Human Error: Self-Initiated Other-Repair. It certainly seems possible that the human would be unsure about the next move in a movement sequence, either through uncertainty about some (possibly implicitly) assumed set of rules or through a memory lapse. There are two ways that this might be visually communicated: a cessation of activity, or the generation of some action but in an 'uncertain', perhaps hesitant manner. Humans are good at detecting uncertainty in others and it seems that a variety of visual cues might be involved in this, perhaps including facial expression and general bodily tension and hesitancy. However, getting a robot to detect uncertainty seems extremely difficult at best. If the robot could detect the human's uncertainty then it could prompt the human in some way e.g. by speaking or by repeating the previous action in hopes of triggering a memory through association.

Human Error: Other-Initiated Other-Repair. In the conversational case this would take the form of the human saying something and the robot producing a corrected version of the human's word or sentence. In the visual case we could argue that the human performing an action other than what the robot expected, with the robot then demonstrating the action the human should have performed, constitutes a roughly equivalent structure. This still leaves us, though, with the problem of how the human is to know that the robot's action is a correction and not simply a response. The robot would have to provide an explicit signal along the lines of the negation behaviour described earlier before demonstrating or describing the correct action.

Robot Error: Self-Initiated Self-Repair. We consider the case where a robot is responding to an action that it has assumed to be c_1 that, as the human continues moving, turns out to be c_2 as an example of the robot catching its own error and thus initiating its own repair. In this case the appropriate behaviour would seem to be to stop, perhaps to engage in some kind of negation behaviour,

[17] We note that a similar strategy is employed by an autistic child in [23].

and then to perform the correct action. Of course, if the human continues to act through all this then the robot should start acting again from the latest action: presumably the human has ignored or missed the repair behaviour (or the original error) if he continues to act through it.

Robot Error: Other-Initiated Self-Repair. For this to work the robot needs to be able to detect the human's initiation, which again leaves us with the difficult problem of detecting and correctly interpreting uncertainty and surprise reactions. In this case, the robot could shift to the next most probable interpretation for the current action, or maintain the present interpretation but shift to a less likely expectation. It should engage in negation behaviour first to make it clear that it is abandoning the previous action and switching to a different one.

Robot Error: Self-Initiated Other-Repair. This seems very useful, as it would cover cases where the robot does not know what action to take. If the robot can signal its uncertainty then the human can show it what to do, for instance demonstrating that the robot should move forward by approaching the robot, or that it should move away by retreating away from the robot.

Robot Error: Other-Initiated Other-Repair. Provided that the robot can detect the human's initiation, this would function very similarly to the self-initiated other-repair case. The robot would have to stop as soon as it detected the human's initiation and observe the demonstrated action, then repeat that action itself. Detection of initiation, as always, is the significant problem here.

3.5 Summing Up

We note that a common problem in the above section is the issue of identifying cases where a problem has occurred — it seems that we can get around this problem in many cases by prespecifying a simple 'negation behaviour'. It seems that by assuming the use of explicit signalling (negation) behaviour by both parties to indicate that repair is being engaged in, repair of a limited kind does become an option even in our simple visual interaction. However, the detection of such negation behaviour seems problematic and it is unclear to what extent negation behaviour is commonly used in human interaction: we do not know if we can expect such behavior or not. Repair is potentially very useful as it gives both parties in the interaction a greater ability to make their expectations clear and to influence the behaviour of the other interactant. In our particular application domain, where we study the interaction of children with autism with a mobile robot [24, 25, 23], repair seems especially important for the children, some of whom already attempt to give the robot verbal instructions as to how it should behave which, of course, it ignores. For these children the ability to correct the robot and to get it to behave in a way that they consider appropriate could be very valuable.

4 Measurement

Our goal in all of this, of course, is to create an agent capable of having a 'successful' interaction with a human, based on criteria used for the measurement of human–human interaction rather than on external optimization criteria. It is difficult to define exactly what constitutes a successful interaction, which partly follows from the difficulty of measuring interaction. There are a number of measures we can consider, though. First, of course, there is the possibility of using CA to analyze our interactions [23], giving us a rigorous qualitative assessment of the success of the interaction. Equally, we can assess the satisfaction of the human with the interaction through interviews and questionnaires. We can also suggest more quantitative measures, such as duration of interaction or time human spends gazing at the robot [23]. We also note the existence of the tool THEME, a statistical program which seeks patterns in movement and which has been used for the study of interactive behavior[26–28]. [4] also provides means of examining the interactive behaviour of virtual agents, albeit in a quite different context. In each case we can use a robot controlled by a simpler, non-interaction-aware controller (e.g. following heat sources, as in the Aurora project [24, 25, 23]) as a control.

5 Conclusion and Other Work

We have considered application of a number of ideas derived from the human sciences in the development of artificial interactive agents. Our initial, brief consideration of heuristics applicable to identification of longer sequences of interaction provides us with the beginning, at least, of a means of constructing a system capable of identifying observed interactive behaviors, albeit in a coarse way. Our more detailed consideration of extension of a simple interactive agent architecture to encompass something closer to 'true' interaction and to incorporate some CA-inspired structures suggests a means of building an agent capable of generating interactive behaviors in a very simple interaction. Hopefully we have demonstrated the usefulness of considering the design of interactive agents in the light of knowledge from the human sciences. We also do not wish to give the impression that we are only interested in the dancing interaction: we are merely using this as an example. Future work will create agents with an architecture based on section 3 of this paper and study their interactions with both humans and other interactive agents. At the moment the agents in question are Khepera robots [29] operating in the Webots simulation environment [30]: the controllers involved will also be ported to real Kheperas.

It seems to us that the sequences of interaction described by CA may usefully form the basis for computational modeling and robot controllers if context is appreciated and the diversity of possible initiations and responses considered, with the action produced always being a matter of local activity and contingencies. At this point we are only dealing with a very limited part of the full complexity of human interaction as described by CA, but we believe that our approach represents a reasonable starting point.

6 Acknowledgements

This work is partially supported by EPSRC grant GR/M62648. We would like to thank Paul Dickerson and John Rae for discussions and explanations about conversation analysis and for their involvement in the Aurora project.

References

1. Maturana, H.R., Varela, F.J.: Autopoiesis and Cognition. D. Reidel, Dordrecht, Holland (1980)
2. Dautenhahn, K., Ogden, B., Quick, T.: From embodied to socially embedded agents: Implications for interaction among robots. Cognitive Systems Research Journal (to appear) Special Issue on Situated and Embodied Cognition.
3. Clark, H.H.: Managing problems in speaking. Speech Communication **15** (1994) 243–250
4. di Paolo, E.: Behavioral coordination, structural congruence and entrainment in a simulation of acoustically coupled agents. Adaptive Behavior **8** (2000) 27–48
5. Yoon, S.Y., Burke, R.C., Blumberg, B.M., Schneider, G.: Interactive training for synthetic characters. In: Proceedings of the Seventeenth National Conference on Artificial Intelligence and Twelfth Conference on Innovative Applications of Artificial Intelligence, Austin, Texas, USA, AAAI Press / The MIT Press (2000) 249–254
6. Cassell, J., Sullivan, J., Prevost, S., Churchill, E., eds.: Embodied Conversational Agents. MIT Press (2000)
7. Kendon, A.: A description of some human greetings. [8]
8. Kendon, A.: Conducting Interaction: Patterns of Behavior in Focused Encounters. Cambridge University Press, Cambridge, UK (1990)
9. Kendon, A.: Features of the structural analysis of human communicational behavior. In von Raffler-Engel, W., ed.: Aspects of Nonverbal Communication. Swets and Zeitlinger, Lisse, Netherlands (1980)
10. Schank, R.C., R, A.: Scripts, plans, goals and understanding. Lawrence Erlbaum Associates Inc, Hillsdale, NJ (1977)
11. Collett, P.: Mossi salutations. Semiotica **45** (1983) 191–248
12. Levinson, S.C.: Pragmatics. Cambridge University Press, Cambridge, UK (1989)
13. Sacks, H., Schegloff, E.A., Jefferson, G.: A simplest systematics for for the organization of turn-taking for conversation. Language **50** (1974) 696–735
14. Hutchby, I., Wooffitt, R.: Conversation Analysis. Polity Press, Cambridge, UK (1998)
15. Dautenhahn, K.: Embodiment and interaction in socially intelligent life-like agents. In Nehaniv, C., ed.: Computation for Metaphors, Analogy and Agents. Springer-Verlag, Berlin, Germany (1999)
16. Luff, P., Gilbert, N., Frohlich, D., eds.: Computers and Conversation. Academic Press Ltd, London, UK (1990)
17. (Ferrell), C.B.: Early experiments using motivations to regulate human-robot interaction. In: Proceedings of 1998 AAAI Fall Symposium: Emotional and Intelligent, The Tangled Knot of Cognition, Orlando, FL, USA. (1998)
18. Breazeal, C.: Sociable machines: Expressive social exchange between humans and robots. PhD thesis, Department of Electrical Engineering and Computer Science, MIT (2000)

19. Breazeal, C., Scassellati, B.: Infant-like social interactions between a robot and a human caretaker. Adaptive Behavior **8** (2000) 49–74
20. Bryan, K., McIntosh, J., Brown, D.: Extending conversation analysis to non-verbal communication. Aphasiology **12** (1998) 178–188
21. Robinson, J.D.: Getting down to business: Talk, gaze and body orientation during openings of doctor-patient consultations. Human Communication Research **25** (1998) 97–123
22. Frohlich, D., Luff, P.: Applying the technology of conversation to the technology for conversation. [16] chapter 9
23. Dautenhahn, K., Werry, I., Rae, J., Dickerson, P., Stribling, P., Ogden, B.: Robotic playmates: Analysing interactive competencies of children with autism playing with a mobile robot. In Dautenhahn, K., Bond, A., Cañamero, L., Edmonds, B., eds.: Socially Intelligent Agents - Creating Relationships with Computers and Robots. Kluwer Academic Publishers (to appear)
24. Werry, I., Dautenhahn, K.: Applying mobile robot technology to the rehabilitation of autistic children. In: Proceedings of the Symposium on Intelligent Robotic Systems (SIRS'99), Coimbra, Portugal (1999) 265–272
25. AURORA: http://www.aurora-project.com. Last referenced 21st January 2002 (2002)
26. Magnusson, M.S.: Hidden real-time patterns in intra- and inter-individual behavior: Description and detection. European Journal of Psychological Assessment **12** (1996) 112–123
27. Tardif, C., Plumet, M.H., Beaudichon, J., Waller, D., Bouvard, M., Leboyer, M.: Micro-analysis of social interactions between autistic children and normal adults in semi-structured play situations. International Journal of Behavioral Development **18** (1995) 727–747
28. Grammer, K., Kruck, K.B., Magnusson, M.S.: The courtship dance: Patterns of nonverbal synchronization in opposite-sex encounters. Journal of Nonverbal Behavior **22** (1998) 3–29
29. K-Team: http://www.k-team.com. Last referenced 21st January 2002 (2002)
30. Cyberbotics: http://www.cyberbotics.com. Last referenced 21st January 2002 (2002)

A Some CA Notation

The following is adapted from [14].

CAPITALS	Speech is noticeably louder than that surrounding it.
!	Exclamation marks are used to indicate an animated or emphatic tone.
.	A stopping fall in tone - not necessarily the end of a sentence.
?	A rising inflection - not necessarily a question.
<u>Under</u>	Underlined fragments indicate speaker emphasis.
(n)	The number in brackets indicates a time gap in tenths of a second.
(( ))	A nonverbal activity e.g. ((banging sound)), or a transcriber's comment.
[	Indicates the end of a spate of overlapping talk.